

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

L'IMPACT DE L'ACTIVITÉ SPORTIVE SUR LES RÉSULTATS
SCOLAIRES

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE

PAR

TAMARA CAP

SEPTEMBRE 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Tout d'abord, je tiens à remercier M. Philip Merrigan, mon directeur de mémoire, pour sa patience, son support et ses encouragements depuis le début de ce travail de recherche. J'aimerais aussi remercier M. Pierre Lefebvre, codirecteur, qui fût d'une aide cruciale. Ensuite, je me dois de saluer et remercier M. Frank Larouche, analyste du centre de recherche interuniversitaire québécois de statistiques sociales de Montréal, pour tout son aide, son temps, sa patience et sa compréhension au cours des deux dernières années et M. Frédéric Broussau, analyste du centre de recherche interuniversitaire québécois de statistiques sociales de Montréal, pour son aide indispensable dans la construction des fondations du programme pour ce travail. Finalement, merci aux membres du CIQSS pour le soutien financier à travers leur programme de bourses pour une Bourse d'accueil.

Les analyses contenues dans ce texte ont été réalisées au Centre interuniversitaire québécois de statistiques sociales (CIQSS), membre du Réseau canadien des centres de données de recherche (RCCDR). Les activités du CIQSS sont rendues possibles grâce à l'appui financier du CRSHC, des IRSC, de la FCI, de Statistique Canada, du FRQSC ainsi que de l'ensemble des universités québécoises qui participent à leur financement. Les idées exprimées dans ce texte sont celles de l'auteur et non celles des partenaires financiers.

DÉDICACE

À mes parents, ma sœur et mon entraîneur Alfredo.

Merci pour votre support constant depuis le premier jour où j'ai entrepris ce périple. Plusieurs hauts, plusieurs bas, mais enfin le bout du tunnel.

TABLE DES MATIÈRES

LISTE DES TABLEAUX.....	vi
RÉSUMÉ	vii
INTRODUCTION	1
CHAPITRE I	
DONNÉES.....	9
1.1 Échantillon initial	9
1.2 Échantillon final	12
MÉTHODOLOGIE.....	23
CHAPITRE II	
MODÈLES.....	26
2.1 Modèles de base.....	27
2.1.1 Modèle d'estimation avec effets aléatoires	27
2.1.2 Modèle d'estimation à effets fixes	28
2.1.3 Modèle de régression linéaire par moindres carrés ordinaires	29
2.2 Modèles modifiés	30
CHAPITRE III	
RÉSULTATS.....	31
3.1 Résultats modèles de base	31
3.1.1 Résultats : Modèle à effets aléatoires	31
3.1.2 Résultats : Modèle à effets fixes.....	36
3.1.3 Résultats : Test Hausman	41
3.1.4 Résultats : Modèle de régression linéaire	42

3.2	Résultats : Modèles modifiés	44
3.2.1	Résultats : Modèle à effets aléatoires	44
3.2.2	Résultats : Modèle à effets fixes	46
3.2.3	Résultats : Test Hausman	49
3.2.4	Résultats : Modèle de régression linéaire	49
CHAPITRE IV		
	DISCUSSION DES RÉSULTATS	52
	CONCLUSION	58
	APPENDICE A	60
	BIBLIOGRAPHIE	66

LISTE DES TABLEAUX

Tableau	Page
TABLEAU 1: ÉCHANTILLONNAGE DE L'ELNEJ : ÂGE À CHAQUE CYCLE.....	11
TABLEAU 1.1: CROISÉ : ÂGE ET CYCLE.....	12
TABLEAU 1.2: STATUT SPORTIF SELON L'ÂGE DE L'ENFANT	13
TABLEAU 1.3: STATUT SPORTIF SELON LE SEXE DE L'ENFANT	14
TABLEAU 1.4: MOYENNE DU SCORE À L'EXAMEN DE MATHÉMATIQUES	14
TABLEAU 1.5: LISTE DES VARIABLES	20
TABLEAU 1.6: TAUX DE RÉPONSE À L'ELNEJ PAR PROVINCE	22
TABLEAU 3.1: ESTIMATEURS: RÉGRESSION À EFFETS ALÉATOIRES	32
TABLEAU 3.2: ESTIMATEURS: RÉGRESSION À EFFETS FIXES	38
TABLEAU 3.3: ESTIMATEURS: RÉGRESSION LINÉAIRE	43
TABLEAU 3.4: ESTIMATEURS: RÉGRESSION À EFFETS ALÉATOIRES - ACTIFS	45
TABLEAU 3.5: ESTIMATEURS: RÉGRESSION À EFFETS FIXES - ACTIFS	48
TABLEAU 3.6: ESTIMATEURS: RÉGRESSION LINÉAIRE - ACTIFS	51
TABLEAU 4.1: DONNÉES ÉQUATIONS (2).....	57
TABLEAU A.1: BREUSH-PAGAN: XTREG, RE - SPORTIF.....	60
TABLEAU A.2: BREUSH-PAGAN: XTREG, RE - ACTIF	61
TABLEAU A.3: TESTS VARIABLES EXPLICATIVES PRINCIPALES.....	61
TABLEAU A.4: HAUSMAN - SPORTIF.....	62
TABLEAU A.5: HAUSMAN - ACTIF	63
TABLEAU B.1: STATISTIQUES DESCRIPTIVES - SPORTIFS	64
TABLEAU A.5: STATISTIQUES DESCRIPTIVES - ACTIFS.....	65

RÉSUMÉ

Le travail de recherche présenté ici constitue une exigence du programme de maîtrise en économique de l'Université du Québec à Montréal, entrepris en août 2013, sous la supervision de M. Philip Merrigan et M. Pierre Lefebvre.

L'objectif principal de ce mémoire est de déterminer si le fait de pratiquer une activité sportive hors de l'établissement scolaire plus d'une fois par semaine a un impact significatif sur les résultats à un test de mathématiques, soit le CAT/2, qui sera la variable dépendante des modèles de régression pour mesurer la performance scolaire des étudiants observés. Les résultats, bien que faibles, corroborent cette hypothèse.

La question sous-jacente à cette étude, la conciliation entre le sport et les études, ainsi que les bienfaits et les conséquences qui y sont reliés, a été le sujet de plusieurs études commentées dans la revue de la littérature.

À l'aide de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) un échantillon de 1300 étudiants âgés de 7 et 9 ans et observés deux fois du premier au troisième cycle est construit afin d'estimer des modèles de régression pour des données de panel.

Trois méthodes d'estimation sont utilisées pour estimer l'impact du sport sur les performances scolaires : un modèle de régression à effets fixes, un modèle de régression à effets aléatoires et un modèle de régression simple par moindres carrés ordinaires.

Mots-Clés : panel, longitudinal, sport, physical activity, academic performance.

INTRODUCTION

Le ministère de l'Éducation, du Sport et du Loisir (MELS) s'efforce à chaque année à promouvoir le sport, la réussite scolaire et la santé chez les étudiants du primaire jusqu'au secondaire sur tout le territoire québécois. Il met à la disposition des communautés plusieurs associations sportives qui permettent aux jeunes de pratiquer leur sport. Tout récemment, il y a eu l'implantation d'un crédit d'impôt pour les activités physiques des enfants des contribuables afin d'inciter les parents à motiver leurs enfants à pratiquer un sport ou une activité parascolaire. De plus, le ministère a augmenté le nombre d'heures dédié aux cours d'éducation physique dans les établissements du Québec. Tous ces gestes que pose le MELS ne sont pas sans motivation.

Cette thèse cherche à démontrer si l'activité sportive a un impact sur les résultats scolaires et si ce lien de causalité est significativement important. Le développement durable est devenu un sujet populaire au cours des dernières années, mais il faut aussi tenir compte du développement des humains qui bâtiront notre société du futur. Il est donc essentiel pour la société d'être au courant de l'effet de la pratique sportive sur le rendement académique d'un étudiant, car elle pourrait utiliser le sport pour contrer le décrochage scolaire, pour demander un meilleur financement du sport à ses élus et permettre au Québec de mieux se classer au sein des pays de l'OCDE (MELS, 2010a).

L'étude la plus similaire à notre travail de recherche est celle de Shepard *et al.*, originalement écrit en 1984. Au total, 546 étudiants d'écoles primaires de Trois-Rivières ont été retenus. Les établissements qui ont pris part à cette étude ont divisé leurs étudiants en deux groupes. En retranchant du temps des autres matières, sauf les cours d'anglais, le groupe expérimental a pu bénéficier de 5 heures d'éducation physique par semaine. Les autres étudiants recevaient le minimum de 40 minutes

requis par le ministère de l'Éducation du Québec. Ces deux groupes ont été suivi de la première à la sixième année (Shepard, 1997). De plus, ils ont pu compiler des résultats à l'aide de tests sur l'intelligence et sur le développement psychomoteur. Les tests utilisés sont le « Goodenough » test et le WISC test (Wechsler Intelligence scale for Children). À première vue, les résultats du WISC test démontrent que les étudiants du groupe expérimental ont obtenu des résultats supérieurs à ceux des étudiants des autres groupes. Les auteurs ont voulu vérifier l'impact d'une augmentation du nombre d'heures allouées à l'éducation physique sur les résultats en français, en mathématiques, en science de la nature et sur le comportement. Ces résultats ont été compilés à chaque année pour chaque étudiant participant. Pour évaluer l'impact du programme expérimental sur les résultats académiques et pour présenter une comparaison entre les résultats obtenus par les deux groupes, les auteurs se sont servis du programme MANOVA, soit une technique d'analyse de variance avec plusieurs variables dépendantes qui permet ainsi de tester la différence entre plusieurs vecteurs de moyennes. En première année, les résultats semblent démontrer que les étudiants du groupe non-expérimental ont obtenu, en moyenne, de meilleurs résultats que le groupe expérimental. Cependant, cette tendance est renversée pour toutes les autres années, surtout en deuxième, troisième, cinquième et sixième année. En mathématiques, les groupes expérimentaux ont obtenu de meilleurs scores dans quatre comparaisons et des résultats égaux aux autres groupes dans huit comparaisons. Le groupe expérimental n'a pas obtenu de résultats plus faibles et ce, pour toutes les comparaisons. Cette étude longitudinale présente des résultats qui corroborent les hypothèses de ce mémoire. Ici aussi, des facteurs socio-économiques et démographiques, l'âge et le sexe ont été inclus dans l'analyse. Cependant, plusieurs points importants différencient les travaux. D'abord, la méthodologie utilisée est une régression longitudinale alors que l'expérience de Shepard *et al.* utilise une méthode de comparaison. De plus, l'étude de Trois-Rivières a été menée dans cette ville, soit une ville majoritairement francophone. L'ELNEJ offre un échantillon plus vaste et diversifié.

Le texte de Carlson *et al.* (Carlson, 2007) s'appuie sur l'enquête longitudinale « *The Early Childhood Longitudinal Study, Kindergarten Class of 1998 to 1999* » qui est un échantillon représentant des enfants entrant la prématernelle (Kindergarten) en 1998 jusqu'à leur cinquième année de primaire en 2004. Cette analyse examine l'influence de l'activité physique sur les résultats en mathématique et en lecture à l'aide d'un échantillon de 5316 étudiants représentant une étude longitudinale menée de 1998 à 2004 dans des écoles primaires américaines. En tenant compte du revenu familial, de l'ethnie, de l'éducation de la mère, du sexe de l'enfant ainsi que de plusieurs variables démographiques, les auteurs ont effectué une régression linéaire multivariée longitudinale à l'aide du programme SUDAAN pour tester la relation entre l'activité physique et les scores en mathématiques et en lecture. À chaque année, les résultats dans ces deux domaines ont été compilés en contrôlant pour le niveau de scolarité de l'enfant. De façon générale, les résultats réfutent leur hypothèse initiale. Chez les filles, seuls les résultats en lecture des étudiantes de cinquième année étaient significatifs et positivement reliés à la pratique d'activités physiques. Chez les garçons, aucune différence n'a été notée dans les résultats en mathématiques et en lecture. Les résultats négatifs obtenus dans cette enquête peuvent être liés à la façon dont les données ont été manipulées afin de produire un échantillon représentatif de la population américaine.

L'étude canadienne « *The relationship between physical activity, self-esteem, and academic achievement on 12 year-old children* », a été menée par Mark S. Tremblay *et al.* (Tremblay, 2000). Ils ont tenté d'étudier la relation entre la participation à des activités physiques, l'estime de soi, un index de poids et les performances académiques pour des étudiants en sixième année de primaire au Nouveau-Brunswick. Tout comme pour ce mémoire, les auteurs ont pris le soin de tenir compte de certains facteurs socio-économiques. Au total, 74,3 % de la population totale d'étudiants de sixième année du Nouveau-Brunswick ont répondu au questionnaire

qui leur a été envoyé afin de recueillir les informations concernant le sexe, l'âge, la structure familiale et le statut socio-économique des familles. Quatre questions ont été utilisées pour différencier les niveaux d'activité physique, soit: combien de fois dans la dernière semaine l'enfant a-t-il pratiqué un sport où le niveau d'intensité cardiorespiratoire était élevé, combien de fois dans la dernière semaine l'enfant a-t-il fait du vélo ou de la marche durant au moins 20 minutes, combien de fois dans la dernière semaine l'enfant a-t-il fait des exercices d'étirement et combien de fois dans la dernière semaine l'enfant a-t-il fait des exercices de musculation (Tremblay, 2000). En régressant le résultat au test de mathématiques sur les variables d'activité physique, soit un agrégat pondéré des résultats aux quatre questions mentionnées plus haut, le sexe, le statut socio-économique, la structure familiale, le nombre de frères et sœurs et l'estime de soi, les auteurs ont trouvé une relation négative significative entre les résultats en mathématiques et l'activité physique. Ces résultats peuvent être expliqués par le fait que la population est représentative du Nouveau-Brunswick seulement. On ne peut donc pas extrapoler au Canada. De plus, la courte période sur laquelle les données ont été collectées et le peu de variables explicatives peuvent être responsables de biais dû à la présence d'hétérogénéité ou de corrélation entre les variables explicatives. Si l'hypothèse d'homogénéité n'est pas respectée, les estimés de la régression seront biaisés et ne pourront être utilisés pour des fins d'inférences. Ces types de problèmes ne risquent pas d'apparaître dans ce mémoire étant donnée le plus grand nombre de variables explicatives et que le panel s'étend sur plusieurs périodes. De plus, comme ils le soulignent dans leur conclusion, les questions posées aux enfants pour mesurer le niveau d'activité physique manquent de précision. Celles de l'ELNEJ permettent de séparer en cinq niveaux la fréquence à laquelle l'enfant pratique le sport. Finalement, la variable utilisée pour mesurer la performance scolaire dans ce mémoire est un test standardisé à l'échelle nationale et administré à tous les étudiants durant les huit cycles de l'ELNEJ tandis que dans le texte de Tremblay *et al.*, le test n'est administré qu'aux étudiants de sixième année au Nouveau-Brunswick.

L'article d'Elizabeth Mulholland « L'influence du sport : le rapport sport pur. » (Mulholland, 2008) propose des conclusions qui consolident celles des études présentées plus haut. Suite à la pratique d'une activité physique, l'enfant est plus détendu ce qui lui permet d'améliorer sa mémoire et d'être plus concentré. La résolution de problèmes se fait avec plus de latitude étant donné la meilleure concentration.

Le texte de Mulholland se base, entre autre, sur un article souvent cité dans la littérature, soit celui de Richard Bailey dans *Journal of School Health* « *Physical Education and Sport in Schools : A Review of Benefits and Outcomes* » (Bailey, 2006). Dans son texte, l'auteur fait une revue des essais scientifiques qui ont analysés les effets du sport sur les résultats scolaires. Il se base sur plusieurs points de son inspection des articles pour finalement porter une conclusion qui appuie la prémisse présentée au début de ce travail. Ainsi Bailey conclue que :

Overall, the available research evidence suggests that increased levels of physical activity in school—such as through increasing the amount of time dedicated to PES—does not interfere with pupils' achievement in other subjects (although the time available for these subjects is consequently reduced) and in many instances is associated with improved academic performance (Bailey, 2006).

Le texte *Évaluation de l'expérience des participants au programme sport-études* proposé par le Ministère de l'Éducation, du Loisir et du Sport (MELS, 2013) est un croquis partiel de ce mémoire. Ce document présente un sommaire des résultats

académiques obtenus par des étudiants inscrits au programme de sport-études versus des élèves qui ne le sont pas. Il fait aussi état de l'arbitrage que doivent faire ces jeunes entre le temps dédié au sport et celui consacré aux études et ce, selon le point de vue des étudiants-athlètes, de leurs entraîneurs et de leurs parents. Les résultats obtenus dans l'étude menée par le MELS sont significatifs: les étudiants inscrits au programme sport-études ont de meilleurs résultats en Histoire, en Mathématiques et en Français.

Le résumé d'Anne Guèvremont *et al.* « Une forte participation à des activités parascolaires contribue-t-elle ou nuit-elle au développement de l'enfant? » (Statistique Canada, 2010) pose plusieurs constats qui s'alignent avec les attentes de ce mémoire. Parmi les points importants, il semblerait que les activités parascolaires aient une influence positive, mais limitée, sur les résultats académiques, plus précisément en mathématiques. De plus, le lien de causalité entre la pratique d'activités sportives et les rendements académiques est plus significatif pour les activités exercées hors de l'établissement scolaire que celles réalisées à l'école. Cependant, ce constat ne tient que pour le groupe d'âge des 14-17 ans.

Dans un autre ordre d'idée, le livre de Francine Lauzon, *L'éducation psychomotrice : Source d'autonomie et de dynamisme* (Lauzon, 1990) présente les éléments essentiels au développement d'un enfant en regroupant des aspects cognitifs, socio-affectifs et psychomoteurs. Elle consacre un chapitre à l'importance et l'impact de l'activité physique sur le développement des jeunes enfants. Ses conclusions sont claires quant au développement de l'autonomie, la confiance en soi et de l'estime de soi. L'atteinte de ces objectifs passe par la conquête d'habiletés motrices et sensorielles qui s'amplifient avec la pratique d'activité physique. L'auteure affirme aussi que les activités sensorielles et motrices sont nécessaires au développement de l'intelligence et du rapport socio-affectif de l'enfant.

Le chapitre « Le continu du discours : les prolongements éducatifs » tiré du livre *Sport et délinquance* de Jean-Yves Lassalle (1988) suggère des conclusions au sujet de la réhabilitation des délinquants à travers le sport. Ses conclusions peuvent être mises en parallèle avec cet essai, car elles portent sur les dimensions psychologiques et relationnelles d'un individu. Sur le plan psychologique, il souligne que l'activité sportive permet de développer des qualités comme le respect des règles, la maîtrise de soi et la satisfaction personnelle qui sont des éléments nécessaires pour le bien-être d'un individu. Du point de vue relationnel, l'auteur atteste que le sport, et surtout le sport d'équipe, permet à un individu de mieux exprimer ses sentiments, de développer un sentiment d'appartenance et est un excellent canal pour introduire le travail d'équipe.

Finalement, il existe des études qui se sont lancées sur les mêmes pistes que celles présentées plus haut, mais avec des résultats différents. Dans celle de James F. Sallis *et al.* « *Effects of Health-Related Physical Education on Academic Achievement : Project SPARK* » (Sallis, 1999), les auteurs présentent une étude qu'ils ont menée pendant deux ans auprès de 7 écoles primaires dans une banlieue du sud de la Californie. Les variables utilisées pour mesurer les résultats scolaires sont les tests du Metropolitan Achievement Test, MAT6 et MAT7, soit un test de mathématique et un test de psychologie. Ils cherchaient à démontrer si une augmentation des heures dédiées aux cours d'éducation physique serait reliée à une augmentation des performances académiques. Leur constat fut que ce n'était pas le cas. Il faut tenir compte du fait que l'échantillon utilisé pour leur étude ainsi que la durée du panel sont de taille faible et assez court, ce qui cause un problème de non-stationnarité des données, comme mentionnés dans la section Méthodologie. L'enquête utilisée pour la présente étude ne fait pas face à ce genre de problème.

La problématique qui entoure l'étude de l'effet du sport sur les résultats scolaires repose sur plusieurs aspects : le sport, bien entendu, mais aussi le développement humain, le financement et l'importance que les Québécois accordent au sport. Il faudra s'intéresser à plusieurs domaines afin de bien saisir l'amplitude des bienfaits du sport sur les rendements académiques.

À la lumière de ces informations, cette recherche tentera de démontrer qu'au-delà de l'agressivité et de l'instinct de compétition qui peuvent être développés en pratiquant certains sports, les bienfaits qu'a la pratique d'activités sportives sur les résultats scolaires peuvent être significativement importants. Certes, ces comportements négatifs sont peu souhaitables et la violence dans le sport est un phénomène préoccupant qui a pris de l'ampleur au cours des dernières années, mais il ne faut pas oublier que ce travail porte sur des étudiants sous la responsabilité de parents, enseignants et entraîneurs qui ont le devoir d'apprendre à ces jeunes comment utiliser le sport comme antidote à de tels sentiments. Ainsi, avec des conclusions qui appuient les postulats présentés ci-haut, ces résultats permettront d'éclairer, en partie, la tâche de notre société qui est en charge d'éduquer la génération des adultes de demain.

CHAPITRE I

DONNÉES

1.1 Échantillon initial

L'objectif de ce travail est d'établir un lien empirique entre l'activité sportive et la performance à l'école avec l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) de Statistique Canada, qui est biennale et qui couvre les années 1994-1995 à 2008-2009. Le site web de Statistique Canada la décrit ainsi :

L'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) est une étude à long terme sur les enfants canadiens qui permet de suivre leur croissance et leur bien-être de la naissance au début de l'âge adulte. Ayant débutée en 1994, l'ELNEJ est menée conjointement par Statistique Canada et Ressources humaines et Développement des compétences Canada (RHDC), connu auparavant sous le nom de Développement des ressources humaines Canada (DRHC). L'enquête a été conçue pour recueillir des renseignements sur les facteurs qui influent sur le développement social et émotionnel ainsi que sur le comportement des enfants et des jeunes. Elle permet en outre de suivre les conséquences de ces facteurs sur leur développement au fil du temps. L'enquête porte sur un large éventail de sujets y compris la santé, le développement physique, l'apprentissage et le comportement des enfants ainsi que des données sur leur environnement social (famille, amis, école et communauté).

La variable utilisée pour mesurer la performance scolaire est le score d'un test en mathématiques, le CAT/2, dont la difficulté augmente avec le niveau de l'année scolaire. Ce test est produit à tous les cycles (8) de l'ELNEJ pour les enfants à l'école et qui ont de 7 à 15 ans. Ce test mesure les capacités de l'élève à effectuer des opérations d'addition, de soustraction, de multiplication, de division et à résoudre des problèmes comportant des pourcentages. Seul le score gradué est retenu pour ce mémoire (ELNEJ, 2010).

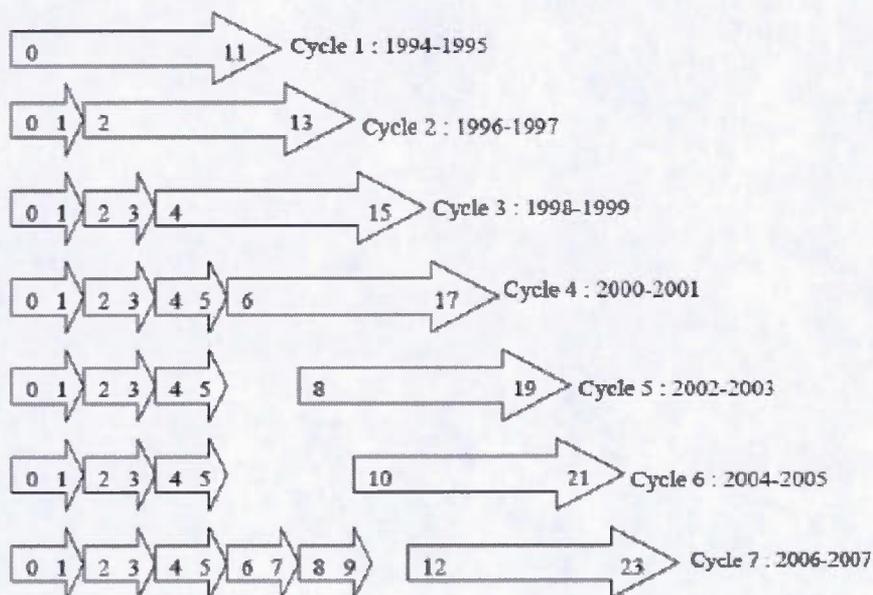
Deux variables indépendantes vont être utilisées pour calculer l'impact de l'activité physique sur les résultats en mathématiques. Plusieurs questions relatives au sport sont disponibles dans l'ELNEJ, mais pour une grande majorité, elles ne sont pas récurrentes à travers les cycles. La variable *actif* représente les étudiants qui pratiquent une activité parascolaire quelconque, alors que la variable *sportif* représente les étudiants qui pratiquent une activité sportive. Pour la variable *sportif*, on demande au répondant : « Dans les 12 derniers mois, en dehors des heures de classe, à quelle fréquence a-t-il/elle participé à des activités physiques ou à des sports libres? ». Les réponses retenues sont : presque tous les jours, quelques fois par semaine, environ une fois par semaine, environ une fois par mois et presque jamais. L'étudiant pour qui le parent a répondu une des trois premières réponses est considéré sportif. Celui dont le répondant a choisi une des deux dernières réponses est classé parmi les non-sportifs. En ce qui a trait à la variable *actif*, la question posée est: « Dans les 12 derniers mois, en dehors des heures de classe, à quelle fréquence a-t-il/elle suivi des leçons ou des cours de musique, de danse ou d'art ou participé à d'autres activités non sportives? ». Le choix de réponses et le processus de catégorisation est le même que pour la variable relative au sport.

La corrélation entre ces deux variables est de 19,77 % dans l'échantillon de régression. La corrélation entre la variable *sportif* et *scoreMath* est de 9,61 % alors qu'elle est de 5,73 % entre *actif* et *scoreMath*. Ainsi, nous pourrions produire des

estimations de l'effet de l'activité physique sur les scores de mathématiques avec des techniques de régression pour données longitudinales. Plusieurs milliers d'enfants ont participé à cette enquête ce qui la rend très attrayante.

L'échantillon initial est constitué d'enfants répartis à travers le Canada âgés entre 0 et 11 ans. La distribution des étudiants à travers les cycles est décrite dans le diagramme suivant où les longues flèches représentent la cohorte initiale alors que les petites représentent les cohortes du Développement de la petite enfance.

Tableau 1. Échantillonnage de l'ELNEJ: Âge à chaque cycle



Source : Statistique Canada, ELNEJ.

Étant donné que le degré de difficulté du test de mathématique, le CAT/2, augmente avec la scolarité, il sera tout aussi important de contrôler l'âge des étudiants afin de pouvoir faire des comparaisons raisonnables. Pour la distribution des individus par âge, présenté dans le Tableau 1 plus haut, il est à noter que du cycle 5 au cycle 8, la distribution du nombre de répondants par âge est incomplète, car la cohorte initiale de 0 à 11 ans a vieilli à chaque cycle sans remplacement. Ceci est dû au manque de répondants pour certains groupes d'âge.

De plus, les cycles 4 à 8 ne sont pas retenus surtout dû au problème des données manquantes quant à la question par rapport au sport. En effet, une modification dans la formulation et le codage de la question font qu'aucune observation est notée après le cycle 3 pour cette question. Étant donné le nombre de répondants trop faible pour certains groupes d'âge et une distribution plutôt hétérogène pour certains cycles, seuls les enfants de 7 et 9 ans seront observés et ce, deux fois du premier au troisième cycle inclusivement.

1.2 Échantillon final

L'échantillon final est constitué de 1300 étudiants âgés de 7 et 9 ans observés au cours des trois premiers cycles, plus précisément, 650 enfants observés deux fois dans la régression. La décomposition du panel par cycle est présentée dans le Tableau 1.1 ci-dessous.

Tableau 1.1 Tableau croisé – Âge et cycles

Cohorte	Cycle 1	Cycle 2	Cycle 3
7 ans-Cycle 1	7 ans = 314 obs.		
7 ans-Cycle 2		7 ans = 318 obs.	
9 ans-Cycle 2		9 ans = 332 obs.	
9 ans-Cycle 3			9 ans = 336 obs.

Source : ELNEJ et calculs de l'auteure.

Deux constats sont tirés de la distribution des jeunes qui ont répondu à la question sur le sport. D'abord, le nombre de sportifs est plus important que le nombre de non-sportifs et ce, qu'ils soient observés à 7 ans ou à 9 ans. Ensuite, lorsque ces jeunes sont observés pour la deuxième fois, on note que 24 étudiants passent du statut de non-sportif à sportif. Ces constats sont présentés dans le Tableau 1.2 ci-dessous.

Tableau 1.2 Statut sportif selon l'âge de l'enfant

		Âge de l'enfant		Total
		7 ans	9 ans	
Sportif	0	282	258	540
	1	368	392	760
	Total	650	650	1300

Source : ELNEJ et calculs de l'auteure.

Lorsque la distribution des enfants qui ont répondu à la question sur le sport est observée par sexe, elle montre que les parents des garçons sont moins nombreux que ceux des filles à répondre. Pour ce même groupe, les sportifs représentent plus du double de garçons non-sportifs alors que chez les filles, les deux groupes ont un nombre plutôt égal d'étudiantes. Ces variables sont présentées dans le Tableau 1.3 suivant.

Tableau 1.3 Statut sportif selon le sexe de l'enfant

Sportif	Sexe de l'enfant		Total
	0 (garçon)	1 (fille)	
0	180	322	502
1	464	334	798
Total	644	656	1300

Source : ELNEJ et calculs de l'auteur.

En ce qui à trait à la distribution des moyennes obtenues au test de mathématiques, il faut souligner que ces moyennes ne varient pas de façon significatives lorsque les sexes sont observés séparément pour tous les cycles. Pour tous les cycles, l'écart entre la moyenne des garçons et la moyenne du test, lorsque les sexes sont confondus, n'est pas significatif. Cette conclusion s'applique aussi pour la moyenne du test obtenue par les filles. De plus, la moyenne varie à la baisse à chaque cycle, sauf pour les deux premiers cycles chez les filles. Cette diminution est explicable par le niveau de difficulté du test qui augmente à chaque année et ce, malgré une graduation faite en fonction de l'année scolaire. Pour l'échantillon de la régression, la moyenne au score de mathématique est de 364 points. Le Tableau 1.4 ci-bas présente les moyennes obtenues au test.

Tableau 1.4 Moyenne du score à l'examen de mathématiques

	Moyenne-Garçons	Écart-type	Moyenne-Filles	Écart-type	Moyenne au scoreMath (combinées)
Cycle 1	367.9	66.90	363.9	66.58	363.8
Cycle 2	364.7	66.66	367.6	67.51	
Cycle 3	340.4	60.30	344.4	63.09	

Source : ELNEJ et calculs de l'auteur.

Évidemment, beaucoup de facteurs autres que les activités sportives peuvent avoir un impact sur les résultats au test de maths. L'ELNEJ offre une vaste diversité de variables de contrôles qui pourraient être potentiellement incluses dans une régression cherchant à estimer l'effet de l'activité sportive sur les résultats en mathématiques.

Le revenu des parents et leur niveau d'éducation jouent un rôle important dans l'éducation de l'enfant. En effet, une famille plus riche, de manière générale, peut acheter plus de ressources utilisables pour l'éducation des enfants et elle peut aussi leurs offrir plus facilement un parcours scolaire en milieu privé. Dans l'étude de Finnie et Muller (Finnie, 2008), les auteurs ont trouvé comme résultats que pour les hommes, un revenu parental de 100 000 \$ et plus est associé à une augmentation de 19,1 points de pourcentage de la possibilité de poursuivre des études universitaires par rapport au groupe de revenu de 50 000 \$ à 75 000 \$. Pour les femmes, cette statistique grimpe à 19,8 points de pourcentage.

En ce qui a trait à l'effet du revenu familial indexé de l'indice des prix à la consommation, il a été représenté par quatre variables dichotomiques représentant différents paliers. La distribution des classes de revenu s'avère être plutôt représentative d'une distribution des paliers de revenu d'une population-type. Environ 9 % des répondants ont un revenu élevé, 25 % un revenu moyen-élevé, 47 % un revenu moyen, laissant ainsi 19 % avec un bas revenu, soit sous 30 000 \$.

Dans ce même article, les auteurs divulguent aussi que les garçons ayant des parents avec un diplôme de premier cycle universitaire ont 31,2 points de pourcentage plus de chance de s'inscrire à l'université que ceux ayant des parents qui n'ont pas complétés d'études post-secondaires. Dans ce mémoire, la mère peut avoir obtenu un diplôme d'études secondaires, post-secondaire, collégiales ou universitaires. Sinon, elle est non-répondante. Ces pourcentages sont respectivement 20 %, 31 %, 22 %, 13 % et 14 %. Pour les hommes, seuls les niveaux d'études supérieures sont observés. Dans

cet échantillon, 22 % des enfants ont un père détenant un diplôme d'études collégiales et 13 % un diplôme d'études universitaires. Si ce parcours académique est projeté sur l'étudiant, il est possible de générer plusieurs variables relatives à l'attente des parents quant aux études que complèteront leur enfant. Elles ont été catégorisées en trois variables dichotomiques. Un peu plus de 24 % des répondants estiment que leur enfant complètera des études universitaires, 69 % espèrent des études collégiales laissant ainsi 7 % pour des études secondaires.

De même, la structure familiale est un facteur important dans la réussite scolaire. Par exemple, il sera difficile pour une famille monoparentale de trouver du temps qui permet à l'enfant de participer à des activités sportives structurées. La variable dichotomique *monoparental* prend la valeur de 1 pour un foyer où il n'y a qu'un seul parent et 0 autrement. Un peu moins de 12 % des étudiants observés sont issus de familles monoparentales. Qui plus est, dans l'étude de Deslandes et Cloutier (Deslandes, 2005), les auteurs mentionnent que les enfants vivant dans une famille non traditionnelle ont des résultats académiques plus faibles et présentent plus de problèmes de comportement. La variable *parentsbio* saisit l'effet d'avoir deux parents biologiques dans le foyer, ce qui est le cas pour plus de 82 % des jeunes. Si un des deux n'est pas le parent biologique, cette variable dichotomique prend la valeur de 0. De plus, l'âge de la mère est aussi une caractéristique qui permet d'évaluer le foyer de l'étudiant. Pour 50,07 % des jeunes, la mère est âgée de 25 ans ou moins alors que 45,15 % des enfants ont une mère d'âge moyen, c'est-à-dire entre 26 et 44 ans. Seuls 4,78 % des enfants de l'échantillon ont une mère de 45 ans et plus. Finalement, pour permettre de bien capter l'effet de la structure familiale, trois variables binaires contrôlent le nombre de frères et ou sœurs dans le foyer du jeune étudié. Les enfants ayant seulement un frère ou une sœur sont majoritaires à 51 % dans ce mémoire, à peine 30 % en ont deux alors que 7 % d'entre eux sont dans des familles avec quatre enfants ou plus. Il s'en suit que 12 % des étudiants sont enfants uniques.

Il sera aussi intéressant de retenir le type d'établissement que les élèves fréquentent. Selon le classement 2007-2008 des écoles secondaires au Québec de l'Institut Fraser pour les résultats aux examens provinciaux en quatrième et cinquième secondaire, les dix premières écoles sont des établissements privés. Ainsi, un étudiant qui fréquente une école du secteur privé est susceptible d'obtenir de meilleurs résultats qu'un étudiant du secteur public si ce classement représente bien le niveau académique des étudiants de ces écoles secondaires. Bien que les jeunes observés sont au niveau primaire, cette caractéristique sera tout de même retenue. Quatre variables dichotomiques sont générées. Ainsi, 2,46 % des jeunes fréquentent un établissement privé, 15,38 % une école catholique, 0,38 % tous autres types d'école, laissant 81,78 % des répondants au public. Il sera intéressant de comparer le lien entre les différents types d'établissements scolaires par rapport à l'éducation publique sur les résultats académiques.

Un autre élément qui peut influencer les résultats scolaires d'un enfant est le pays de naissance de ses parents. Cette caractéristique sera captée par la variable binaire, *immigrantfemme*, qui saisit l'effet d'une mère née au Canada ou non. Basé sur les résultats du PISA 2009, l'article de l'OCDE sur l'intégration des immigrants (OCDE, 2013) note, en moyenne, 36 points de moins dans les résultats des tests de compréhension de l'écrit chez les jeunes âgés de 15 ans qui sont nés dans le pays de résidence de parents immigrants en comparaison avec les enfants dont au moins un parent est né dans le pays de résidence. Cette différence s'amplifie au Canada, en Australie, en Israël et aux États-Unis si on contrôle le statut socio-économique familial. Dans cette épreuve, il sera possible, grâce au grand répertoire de variables dans l'ELNEJ, d'estimer l'effet de la province, mais aussi celle de la taille de la région dans laquelle l'élève réside. Pour l'échantillon de la régression, 42 % des étudiants habitent en Ontario, 22 % dans les Prairies, 10 % dans les provinces de l'Atlantique, 13 % pour la Colombie-Britannique et donc 11 % au Québec. En ce qui

à trait aux régions, elles ont été catégorisées en rurale, petite, moyenne et grande, soit toutes des variables dichotomiques. Étant donné la distribution décrite plus haut des jeunes à travers les provinces canadiennes, il est sans surprise que 23,76 % d'entre eux résident dans de petites régions, 11,23 % dans des régions moyennes, 38,92 % dans de grandes régions et 26,09 % dans des régions rurales. Qui plus est, l'étude de l'OCDE offre aussi un classement des résultats en mathématiques par province et classe le Québec, l'Alberta, l'Ontario et la Colombie-Britannique parmi les quatre premiers. Une autre raison pour laquelle la province de résidence constitue une variable importante est relié à l'accès au sport et son financement. Les provinces qui offrent plus de soutien financier, plus de ligues et d'équipes facilitent la pratique du sport. Prenons par exemple le cas de l'athlétisme, en Colombie-Britannique, le soutien financier se fait sous la base de remboursement de dépenses (BC Athletics, 2011). En Ontario (Athletics Ontario, 2013) et au Québec (MELS, 2011), le support est offert sous forme de crédit d'impôt selon des critères de sélection qui sont basés sur les performances de l'athlète.

Dans un même ordre d'idées, la langue aussi est une caractéristique importante lorsqu'il est question de la performance académique d'un enfant. Les enfants se retrouvant dans un milieu académique où l'environnement linguistique est différent de celui à la maison peuvent être désavantagés. D'abord, par le fait qu'avant même de commencer à comprendre la matière enseignée, ils doivent comprendre le langage dans laquelle celle-ci est présentée et ensuite, car il est plus difficile pour des parents qui ne parlent pas la langue d'enseignement d'aider leurs enfants à faire leurs devoirs. Dans la publication de l'OCDE présentée plus haut, on compte en moyenne un écart de 27 points dans les résultats des élèves nés dans le pays de résidence de parents immigrants par rapport aux élèves autochtones lorsque la langue d'enseignement n'est pas celle couramment utilisée à la maison. Pour l'échantillon de la régression, 78,38 % des répondants sont anglophones, 1,8 % bilingue et 19,78 % sont

francophones ou allophones. Ces constats sont cohérents avec la distribution des étudiants à travers les provinces canadiennes.

Certaines caractéristiques ne sont malheureusement pas incluses dans l'ELNEJ. Les caractéristiques telles que la confiance en soi, la persévérance, la volonté et l'esprit de compétition ne sont pas quantifiables bien que toutes ces singularités chez l'enfant ont des impacts importants autant sur la réussite scolaire que sportive. Le support parental, le type et le nombre d'amis ainsi que les habiletés naturelles sont aussi des variables qui ne pourront être incluses dans la régression. La section suivante cherche à qualifier la nature de ces particularités inobservables chez l'enfant et de leurs impacts sur la méthode d'analyse à employer.

Toutefois, l'ELNEJ met à la disposition des chercheurs plusieurs variables particulièrement intéressantes permettant de conduire une recherche avec un grand nombre d'observations. Le Tableau 1.5 présente la liste des variables qui seront retenues pour les modèles évalués.

Description	variables dichotomiques		Moyennes
	0	1	
Statu sportif de l'enfant	non-sportif	sportif	58,4%
statu actif de l'enfant	non-actif	actif	58,4%
sexe de l'enfant	garçon	filles	51,5%
Revenu du ménage: Élevé	bas revenu	revenu élevé	9,1%
Revenu du ménage: Moyen-élevé	bas revenu	revenu moyen-élevé	25,1%
Revenu du ménage: Moyen-faible	bas revenu	revenu moyen-faible	47,6%
Espérance des parents: plus haut niveau d'étude complété par l'enfant: Études collégiales	études secondaires	études collégiales	68,5%
Espérance des parents: plus haut niveau d'étude complété par l'enfant: Études universitaires	études secondaires	études universitaires	24,3%
Mère (ou la femme du foyer) née au Canada	non-immigrante	femme immigrante	7,4%
Âge de la mère (ou la femme du foyer) à la naissance	mère âgée	jeune mère	50,1%
Âge de la mère (ou la femme du foyer) à la naissance	mère âgée	mère d'âge moyen	45,2%
Langue parlée à la maison	français ou autre	anglais	78,4%
Langue parlée à la maison	français ou autre	bilingue	1,8%
Taille de la région	région rurale	petite région	23,8%
Taille de la région	région rurale	région moyenne	11,2%
Taille de la région	région rurale	grande région	38,9%
Province de résidence	Québec	Ontario	42%
Province de résidence	Québec	Les provinces des Prairies	22%
Province de résidence	Québec	Les provinces de l'Atlantique	10%
Province de résidence	Québec	Colombie-Britannique	13%
Plus haut niveau d'étude complété par le père (ou la femme du foyer): Diplôme Secondaire	parent non répondant	Diplôme Secondaire	20,2%
Plus haut niveau d'étude complété par le père (ou la femme du foyer): Études post-secondaires	parent non répondant	Études post-secondaires	31,1%
Plus haut niveau d'étude complété par le père (ou la femme du foyer): Diplôme collégial	parent non répondant	Diplôme collégial	22,6%
Plus haut niveau d'étude complété par le père (ou la femme du foyer): Diplôme universitaire	parent non répondant	Diplôme universitaire	13,1%
Plus haut niveau d'étude complété par le père (ou l'homme du foyer): Diplôme collégial	études secondaires ou moins	Diplôme collégial	22,3%
Plus haut niveau d'étude complété par le père (ou l'homme du foyer): Diplôme universitaire	études secondaires ou moins	Diplôme universitaire	13,5%
Nombre de frères et sœurs	pas de frère ou sœur	1 frère/sœur	51,0%
Nombre de frères et sœurs	pas de frère ou sœur	2 frères/sœurs	30,1%
Nombre de frères et sœurs	pas de frère ou sœur	3 frères/sœurs	7,8%
Type d'école	école publique	école privée	2,4%
Type d'école	école publique	école catholique	15,4%
Type d'école	école publique	autres types d'écoles	0,4%
Type des parents de l'enfant	1 parent non biologique	2 parents biologiques	82,9%
Nombre de parents	2 parents	famille monoparentale	11,9%
Résultat gradué au test en mathématiques	Score/Math		363,84
Âge de l'enfant	Âge		8

Source : ELNEI et calculs de l'auteur

Finalement, le taux de réponse à l'ELNEJ est plus faible que celui qui était espéré. Ces pourcentages sont présentés dans le Tableau 1.6 suivant. Dans le guide qui accompagne le questionnaire de l'ELNEJ, il est expliqué que le taux de réponse est calculé en divisant le nombre d'examens complétés par le nombre d'enfants éligibles. Un faible taux de réponse diminue l'information disponible pour les chercheurs. Aussi, si les caractéristiques des enfants non-répondants ne sont pas captées dans les régressions, elles pourraient influencer le terme d'erreur et causer un biais dans les estimations.

Tableau 1.6 Taux de réponse au questionnaire de l'ELNEJ par province**NLSCY RESPONSE RATES BY PROVINCE**

PROVINCE	SAMPLED ¹⁴ HOUSEHOLDS	RESPONDING HOUSEHOLDS	RESPONSE RATE
Newfoundland	889	803	90.3%
Prince Edward Island	481	422	87.7%
Nova Scotia	1,059	926	87.4%
New Brunswick	980	857	87.4%
Quebec	2,867	2,514	87.7%
Ontario	4,268	3,519	82.5%
Manitoba	1,133	1,001	88.3%
Saskatchewan	1,166	1,039	89.1%
Alberta	1,355	1,213	89.5%
British Columbia	1,381	1,145	82.9%
TOTAL¹⁵	15,579	13,439	86.3%

Source : Guide ELNEJ Cycle 1

MÉTHODOLOGIE

Quelques difficultés se posent pour l'estimation d'un effet non-biaisé des activités sportives sur les résultats en mathématiques. Un premier problème pourrait apparaître dans un contexte de régression, lorsque la participation à de telles activités est corrélée avec des facteurs inobservables qui influent simultanément sur les résultats et l'activité physique. En second, le cas d'un biais endogénéité, soit un problème qui survient lorsque le sens du lien de causalité qui unit le sport et les résultats scolaires est mal défini, provoque aussi des biais dans l'estimation des effets du sport. Il faut donc minimiser l'impact de tels facteurs dans l'analyse de régression de panel.

C'est pour ces raisons que la technique longitudinale qui sera privilégiée pour l'analyse économétrique sera celle qui permet l'inclusion d'effets fixes individuels dans la régression de l'impact des activités sportives sur les résultats en mathématiques. Cette méthode permet d'inclure une constante qui capte tous les effets des autres variables explicatives qui ne varient pas dans le temps comme par exemple, le lieu de naissance de la mère et le type de parent. L'hypothèse de base de cette technique d'analyse est que les entités, dans notre cas, les enfants, sont uniques. Ainsi, on pose comme hypothèse l'exogénéité stricte des variables explicatives conditionnelle à cette constante pour toutes les périodes. Elle permet donc de prévenir les biais dus à une corrélation entre la variable à expliquer et une des variables explicatives inobservées, mais constante dans le temps de manière à interpréter l'effet de l'activité sportive toutes choses étant égales par ailleurs. Un des bémols de cette approche est qu'elle est plus efficace pour calculer des variations intrinsèques que pour calculer l'effet causal des variables explicatives sur la variable à expliquer. La régression à effets fixes n'estimera que les effets pour les variables qui varient, préférablement beaucoup dans le temps. C'est pour cela que la technique d'analyse à effets aléatoires sera aussi utilisée.

Il sera possible de procéder avec la méthode à effets aléatoires pour étudier l'impact des activités sportives, car elle offre des estimateurs consistants et est donc plus efficace que celle à effets fixes (Cameron, 2012). Cette approche permet aux effets non-observés d'être distribués aléatoirement et de spécifier les autres variables qui pourraient potentiellement influencer la variable dépendante. Ainsi, il sera possible de contrôler par exemple le sexe, le revenu, la scolarisation des parents, la taille de la région et la fratrie pour capter l'effet net du sport sur les résultats en mathématique. Ceci donne donc un pouvoir explicatif aux variables, ce qui n'est pas tout à fait le cas avec la méthode à effets fixes. Cette technique d'estimation est basée sur une hypothèse supplémentaire qui nécessite l'absence de corrélation entre l'activité sportive et les effets inobservés. Néanmoins, étant donné l'éventail de variables que l'on retrouve dans l'ELNEJ, il est possible de contrôler assez de facteurs pour probablement éliminer le problème de variables omises. D'ailleurs, l'analyse à effets aléatoires est plus efficace pour faire de l'inférence étant donné que la variance des estimateurs générés est plus petite que celle obtenue pour les coefficients estimés par la méthode à effets fixes (Merrigan, 2014).

En appliquant le test de Hausman, il sera possible de choisir quelle analyse, à effets fixes ou celle à effets aléatoires, est la plus appropriée. Le test de Hausman peut être mené par le logiciel Stata qui sera utilisé pour ce mémoire. Il repose sur le postulat de base du modèle à effets aléatoires, c'est-à-dire que l'activité physique n'est pas corrélée aux facteurs qui ne sont pas inclus dans la régression étant donné que ceux-ci sont aléatoires. Ce test cherche donc à démontrer l'absence de corrélation entre le terme d'erreur et les régresseurs. Ce mode de validation de la technique de régression a pour hypothèse nulle que le modèle à favoriser est celui à effets aléatoires alors que l'hypothèse alternative préconise celui à effets fixes (Torres-Reyna, 2007). Ceci est dû au fait que l'hypothèse d'absence de corrélation entre les variables inobservées et les variables explicatives est levée pour le modèle à effets fixes enlevant du même coup les biais qui lui sont reliés.

Une dernière technique, celle de la méthode de régression linéaire, sera utilisée pour des fins de comparaison avec les méthodes présentées plus haut. Il est possible de prouver avec les variances estimées pour l'erreur et les effets non-observés que le modèle de régression à effets fixes et celui par moindre carré ordinaire sont des cas limites du modèle à effets aléatoires (Merrigan, 2014).

CHAPITRE II

MODÈLES

La modélisation de données de panel trouve ses origines dans le département d'Économie de l'Université Cambridge. Depuis, plusieurs techniques ont été développées, mais deux seront retenues pour ce mémoire et comparées à un modèle de régression linéaire simple. Pour toutes les régressions et les tests effectués dans les sections suivantes, le niveau de significativité a été fixé à 5 %. Tous les modèles présentés reposeront sur la régression linéaire de type :

$$y_{it} = \beta_0 + \beta_1 Z_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} + c_i + u_{it}, \quad (1)$$

où y_{it} est le score au test de mathématique, Z_{1t} est la variable représentant l'activité physique, x_{it} sont les variables explicatives, c_i est une variable non-observée qui ne varie pas dans le temps, soit l'effet fixe ou aléatoire, u_{it} est un terme d'erreur, i est l'identifiant de l'étudiant et t est l'indice de temps.

Pour chaque méthode d'estimation choisie pour calculer l'impact du sport sur les résultats académiques, une version de l'équation (1) modifiée sera utilisée ainsi que les raisons motivant l'utilisation de la technique en question. La variable dépendante est le score au test CAT/2 de mathématique, soit *scoreMath*. Deux variables indépendantes vont être utilisées pour calculer l'impact de l'activité physique sur les résultats en mathématiques. La variable *actif* représente les étudiants qui pratiquent

une activité parascolaire quelconque, alors que la variable *sportif* représente les étudiants qui pratiquent une activité sportive.

2.1 Modèles de base

2.1.1 Modèle d'estimation avec effets aléatoires

Le premier modèle est estimé à partir d'une régression avec estimateurs à effets aléatoires.

Soit l'équation (1A) :

$$y_{it} = \beta_0 + \beta_1 Z_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} + c_i + u_{it}, \quad (1A)$$

où y_{it} est le score au test de mathématique, Z_{1t} est la variable *sportif*, x_{it} sont les toutes variables explicatives retenues jusqu'à présent, c_i est l'effet aléatoire, u_{it} est le terme d'erreur, i est l'identifiant de l'étudiant et t est l'indice de temps.

Deux hypothèses sont requises pour qu'elle soit estimée par la méthode à effets aléatoires. D'abord, on impose l'exogénéité stricte des variables explicatives conditionnelle à la présence de c_i , c'est-à-dire qu'étant donné la présence de l'effet aléatoire, il y a absence de corrélation entre le terme d'erreur et les variables explicatives et ce, peu importe la période. Ensuite, on impose l'absence de corrélation entre les effets aléatoires inobservés et les autres variables explicatives du modèle. Ces deux hypothèses assurent la convergence des estimés. Il sera cependant plus juste d'utiliser l'option « cluster » sur les individus disponible sur Stata pour contrôler l'hétéroscédasticité lors des régressions. Cette méthode d'estimation est appropriée pour ce modèle, car étant donné que le panel est composé de plusieurs périodes et de plusieurs individus, il est possible d'observer les jeunes deux fois et de contrôler plusieurs variables explicatives. De plus, elle permet aux effets inobservés d'être

distribués aléatoirement. Dans le modèle estimé avec Stata, on impose une correction pour l'hétéroscédasticité. L'option « cluster » permet à la régression de panel d'être « robuste » à l'hétéroscédasticité. Seule la puissance des tests sera affectée si ce problème est en fait absent, mais ces mêmes tests restent valides (Merrigan, 2014). Toujours avec ce même logiciel, il est possible de tester si la variance de l'effet aléatoire est nulle grâce au test de Breusch-Pagan de type multiplicateur de Lagrange.

2.1.2 Modèle d'estimation à effets fixes

Le second modèle est estimé avec la méthode d'estimation avec estimateurs à effets fixes.

Soit l'équation (1B) :

$$y_{it} = \beta_0 + \beta_1 Z_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} + D_1 c_1 + D_2 c_2 + \dots + D_n c_n + u_{it}, \quad (1B)$$

où y_{it} est le score au test de mathématique, Z_{1t} est la variable *sportif*, x_{it} sont les variables explicatives, c_i est une variable non-observée qui ne varie pas dans le temps, soit l'effet fixe, D_i est une des variables dichotomiques qui prend la valeur de 1 ou 0, u_{it} est un terme d'erreur, i est l'identifiant de l'étudiant et finalement, t est l'indice de temps.

La présence du terme « $D_n c_n$ » permet de différencier la constante pour chaque individu tout en conservant la même pente. Étant donné que c_n est maintenant une variable à estimer, il est possible de lever l'hypothèse d'absence de corrélation entre l'effet fixe et les autres variables explicatives du modèle ainsi que tous les biais qui y sont reliés. Cette simplification rend la méthode d'estimation à effets fixes plus adéquate pour ce modèle, car malgré les nombreuses variables explicatives, certaines variables, telles la motivation et la persévérance, ne font pas parties de l'étude. Pourtant, ces dernières ont tout de même un impact sur le score en mathématique, sur

la pratique de l'activité physique et bien d'autres variables. La seule hypothèse du modèle est celle d'exogénéité stricte entre le terme d'erreur et les variables explicatives, conditionnellement à l'effet fixe et ce, pour toutes les périodes. Dans ce modèle, toutes les variables, sauf celles pour l'âge de l'enfant et le score au test de math, sont des variables dichotomiques.

La statistique F calculée lors d'une régression à effet fixe vérifie que tous les effets fixes, c_i , sont égaux à zéro. Il est à noter que le modèle à effets fixes ne peut qu'estimer l'impact des variables qui varient beaucoup dans le temps. Cette faiblesse est notable pour cette étude étant donné que plusieurs variables comme, la province, le type d'école, la scolarité des parents ainsi que la classe de revenu ne varient pas dans le temps, mais peuvent avoir un impact significatif sur les résultats académiques.

Finalement, le test de Hausman permet de comparer les équations (1A) et (1B). Ce dernier sera effectué en dernier lieu pour évaluer les deux méthodes et pour déterminer le modèle idéal dans ce contexte.

2.1.3 Modèle de régression linéaire par moindres carrés ordinaire

Le troisième modèle est celui estimé par une régression linéaire par moindres carrés ordinaire (MCO).

Soit l'équation (1C) :

$$y_i = \beta_0 + \beta_1 Z_1 + \beta_2 x_2 + \dots + \beta_n x_n + u_i, \quad (1C)$$

qui consiste en l'équation (1), mais sans indice de temps. Cette régression est générée pour des fins de comparaison avec les modèles présentés plus haut comme mentionné auparavant. La nature du lien entre ces modèles réside dans la comparaison entre la

variance de l'effet inobservé et celle du terme d'erreur. Cette comparaison est représentée par l'équation (2) dans la section Discussion des résultats.

Il est possible d'effectuer un test d'autocorrélation à l'aide du résidu généré par la méthode de régression linéaire. Ainsi, une régression du terme d'erreur estimé sur ce même terme d'erreur retardé d'une période permettra de confirmer ou d'infirmer la présence d'autocorrélation dans le modèle.

2.2 Modèles modifiés

Les modèles modifiés sont les mêmes que ceux estimés dans la section 2.1 avec comme seule différence majeure que la variable explicative, Z_{1t} , est représentée par la variable *actif*. Les trois modèles, soient le modèle à effets aléatoires, le modèle à effets fixes et le modèle de régression linéaire simple sont estimés et contraints aux mêmes tests. Les équations (1A'), (1B') et (1C') représentent les modèles avec la variable *actif*.

Le but de cette modification est de vérifier si c'est le sport qui influence les résultats ou si le fait de pratiquer une activité parascolaire, peu importe son genre, pourrait aussi avoir un impact sur les tests de mathématique. Comme mentionné dans le chapitre précédent, la corrélation entre ces deux variables est de 0,1977.

CHAPITRE III

RÉSULTATS

1.1 Résultats modèles de base

1.1.1 Résultats : modèle à effets aléatoires

Les résultats de la régression de l'équation (1A) effectuée avec la commande « xtreg, re » et avec l'option « cluster » pour corriger pour l'hétéroscédasticité sont présentés dans le Tableau 3.1 ci-dessous. Pour ce modèle, les coefficients significatifs des variables explicatives sont l'âge de l'enfant, les parents qui espèrent voir leur enfant compléter des études collégiales, le lieu de naissance de la mère, si l'étudiant est bilingue, les grandes régions et toutes les provinces, l'éducation de la mère, un homme ayant complété des études collégiales, l'école privée et l'école catholique. Les autres estimateurs de la régression, entre-autre celui de la variable *sportif*, ne sont pas significatifs. L'échantillon de la régression est composé de 1296 observations, soit 648 individus observés deux fois dans le temps, soit à 7 et à 9 ans. L'écart-type estimée de l'effet aléatoire est de 26,50 et celui du terme d'erreur est de 39,70. Ces statistiques seront discutées plus en détail lors de la comparaison entre le modèle estimé par MCO et celui à effets aléatoires. Le R-Carré pertinent pour les régressions avec estimateurs à effets aléatoires est le R-Carré « total » (overall) généré par Stata. Pour cette régression, il est de 0,5471.

Tableau 3.1 Estimateurs : Régression à effets aléatoires

XTREG, RE- Étudiants sportifs			
Variables	B estimé	Err, robust	P-value
sportif	0.91	3.0	0.77
âge	48.19	1.2	0.00
filie	-0.94	3.1	0.76
revenu élevé	2.27	7.0	0.75
revenu moyen-élevé	2.78	5.0	0.58
revenu moyen-faible	-3.84	4.2	0.37
études collégiales	12.51	5.6	0.03
études universitaires	7.25	6.2	0.24
femme immigrante	13.74	5.6	0.01
jeune mère	-4.16	8.5	0.63
mère d'âge moyen	-7.34	8.5	0.39
anglais	-4.77	5.4	0.38
bilingue	-18.51	9.1	0.04
petite région	1.27	4.5	0.78
région moyenne	-6.63	5.3	0.21
grande région	-9.54	4.2	0.02
Ontario	-24.35	6.7	0.00
Les provinces des Prairies	-27.30	7.4	0.00
Les provinces de l'Atlantique	-25.00	6.4	0.00
Colombie-Britannique	-13.57	7.8	0.08
Diplôme Secondaire	17.15	5.4	0.00
Études post-secondaires	24.36	5.1	0.00
Diplôme collégial	16.04	5.5	0.00
Diplôme universitaire	38.11	6.5	0.00
Diplôme collégial	8.50	3.6	0.02
Diplôme universitaire	2.98	5.2	0.56
1 frère/sœur	5.10	5.4	0.34
2 frères/sœurs	8.20	5.6	0.14
3 frères/sœurs	4.50	7.6	0.55
école privée	23.70	9.5	0.01
école catholique	7.27	4.1	0.08
autres types d'écoles	1.09	6.6	0.87
famille monoparentale	-4.53	7.7	0.56
2 parents biologiques	2.05	6.9	0.77
constante estimée	-26.96	17.2	0.12

Homme {

N= 1296
Nombre de groupes = 2
Observations par groupes = 2
R ² within = 0.7553
R ² between = 0.1672
R ² overall = 0.5471
σ effet aléatoire = 26.5429
σ terme d'erreur = 39.78
Rho (fraction de la variance due à u _i) = 0.3080

Source : ELNEJ et calculs de l'auteur.

Le coefficient de l'impact du sport sur les résultats académiques, soit la variable sportif, estime un impact positif de 0,91 point sur le score, mais il n'est pas significatif. L'âge de l'enfant a un impact significatif de 48,19 points. Ceci est tout à fait normal compte tenu du fait que les connaissances des enfants et le niveau de difficulté des examens augmentent à chaque année et ce, malgré le fait que l'examen auquel sont soumis les étudiants ait été amendé de façon à mieux représenter le niveau scolaire de l'enfant. Pour la variable de sexe qui est égale à zéro pour les garçons et un pour les filles, on note une diminution de 0,94 points sur la note lorsque l'étudiant est une fille.

Dans le cas des variables de revenu, lorsqu'on considère le palier de revenu élevé, l'effet sur la note de mathématique est de 2,27 points par rapport à un enfant vivant dans un foyer à bas revenu. Un impact d'amplitude similaire est estimé pour le revenu moyen-élevé, soit de 2,78 points. Par contre, pour un enfant dont le revenu de la famille est considéré moyen-bas, on note une baisse du résultat au test de math par rapport à un étudiant pour qui le revenu familial est bas. Ce point en particulier sera exploré plus en détail dans les sections qui suivent. Le niveau d'études que les parents espèrent voir leur enfant atteindre aussi influence positivement le résultat académique. L'estimateur qui représente des études collégiales complétées est significatif à 3 % et augmente la note de 12,51 points par rapport à un jeune dont les parents croient qu'il complètera des études primaires ou secondaires. Pour des études universitaires, le coefficient est moins important et n'est pas significatif, soit avec un effet positif de 7,25 points sur le score.

Une autre variable qui permet de décrire le foyer de l'enfant est la variable dichotomique qui représente le statut d'immigrant de la mère, c'est-à-dire, si elle est née au Canada ou non. Cet estimateur est significatif à 1 % et propose une augmentation de 13,74 points de la note au test de mathématique lorsque la mère de

l'enfant n'est pas née au Canada. Ce constat est contraire à ceux présentés dans la revue de la littérature, mais il peut être le résultat d'une meilleure adaptation des enfants des femmes immigrantes au Canada. Toujours au sujet de cette dernière, la variable dichotomique *jeuneMere* correspond aux mères âgées de moins de 25 ans, la variable *avgMere* représente celles entre 25 et 34 ans. Ces deux variables ne sont pas significatives, mais tout de même, pour l'échantillon de la régression, le score diminue de 4,16 points pour un enfant ayant une jeune mère par rapport à la note d'un enfant ayant une mère âgée de 45 ans ou plus. Pour les mères d'âge moyen, la note chute de 7,34 points par rapport à celle d'un étudiant dont la mère est considérée âgée.

L'estimateur de la variable dichotomique représentant les étudiants anglophones n'est pas significatif, néanmoins, il estime une baisse du score au test de mathématique de 4,77 points par rapport à celle d'un enfant francophone ou allophone. Dans le cas où l'étudiant est bilingue, cette variation positive est estimée à 18,51 points. Cet échantillon exprime bien la situation d'un pays comme le Canada où les langues officielles sont le Français et l'Anglais. Il est donc attendu qu'un étudiant qui maîtrise les deux performe mieux.

Parmi les coefficients des variables dichotomiques qui captent la taille de la région dans laquelle l'enfant réside, seul celui pour les grandes régions est significatif. Aussi, seul le coefficient pour les petites régions est positif. De plus, la note augmente de 1,27 points pour un étudiant localisé dans une région de moins de 30 000 habitants par rapport à une région rurale. Le score diminue de 6,63 points pour ceux vivant dans des régions de taille moyenne et de 9,54 points pour ceux habitant une région de 100 000 habitants ou plus. Dans un même ordre d'idée, les estimateurs des variables dichotomiques représentant la province où réside l'enfant sont tous négatifs et significatifs. Par rapport au résultat d'un étudiant qui habite au Québec, on note une diminution de 13,57 points pour les étudiants de la Colombie-Britannique,

24,35 points pour ceux de l'Ontario, 25 points pour ceux des provinces de l'Atlantique et 27,30 points pour les jeunes des provinces des Prairies. Déjà, il est possible de conclure que, dans l'échantillon de cette régression, les étudiants québécois ont de meilleurs résultats au test de mathématique. Cette conclusion s'enlignait avec celle tirée au sujet de la langue que parle l'enfant étant donné que le Québec est la province où les deux langues sont le plus couramment utilisées.

En ce qui a trait à l'éducation de la mère, tous les coefficients sont positifs et significatifs. Le score de l'enfant augmente de 17,15 points pour ceux dont la mère a obtenu un diplôme d'études secondaires en comparaison avec ceux dont la mère est non-répondante. La note de l'enfant grimpe de 24,36 points si la mère a complété des études post-secondaires, légèrement au-dessus de 16 points pour un diplôme collégial et 38,11 points pour un diplôme universitaire. Le niveau de significativité et l'ampleur des effets ne sont pas aussi importants pour l'éducation du père. Le score augmente de 8,50 points pour un étudiant dont le père a un diplôme collégial par rapport à un étudiant dont le père est non-répondant. Pourtant, lorsque le père possède un diplôme universitaire, la note n'augmente que de 2,98 points. Le coefficient pour cette dernière variable n'est pas significatif.

Dans un autre ordre d'idées, la variable *privé* qui décrit le type d'établissement que l'enfant fréquente estime une augmentation de la note de 23,70 points par rapport à celle d'un étudiant du secteur public. Cette variation positive est de 7,27 points pour ceux qui fréquentent une école catholique et de 1,09 points pour les autres types d'établissements scolaires.

Finalement, pour ce qui est question de la composition du foyer, les estimateurs des variables dichotomiques générées pour capter l'effet de la taille de la fratrie ne sont pas significatifs. Néanmoins, le score augmente de 5,10 points si l'enfant a un frère ou une sœur versus un enfant unique. Cette variation est de 8,20 points s'il y a 3

enfants au total dans la famille et 4,50 points s'ils sont 4 enfants ou plus. Pour un enfant dans un foyer avec un seul parent, le coefficient de cette variable dichotomique estime une baisse de la note de 4,53 points par rapport à un enfant dans un foyer avec deux parents. Même si ce coefficient n'est pas significatif, le signe de la variation est cohérent. Finalement, le score augmente de 2,05 points pour un enfant dont les deux parents sont ses parents biologiques à comparer un étudiant dont un des deux parents n'est pas son parent biologique.

Le premier test pour le modèle ci-haut est celui de Breush-Pagan, de type multiplicateur de Lagrange, pour tester l'effet aléatoire. L'hypothèse nulle du test est que la variance de l'effet aléatoire est nulle. Le résultat du test, présenté dans le Tableau A.1 de l'appendice, permet de rejeter l'hypothèse nulle. Le test de Breush-Pagan requiert la normalité pour le terme d'erreur, mais comme mentionné dans la section Méthodologie, étant donné qu'un grand nombre d'observations sont retenues au total pour la régression, environ 1300 jeunes, ce test reste valide.

3.1.2 Résultats : Modèle effets fixes

Les résultats de la régression de l'équation (1B) estimée par Stata sont présentés dans le Tableau 3.2 ci-après. Pour ce modèle, les coefficients des variables explicatives qui sont significatifs sont l'âge de l'enfant, un ménage au revenu élevé, l'âge de la mère et les femmes ayant un diplôme universitaire. Les autres variables de la régression, entre-autre la variable de sport, ne sont pas significatives. Pour la régression à effets fixes, les estimateurs pour les variables relatives au sexe de l'enfant et si la mère est née au Canada ou non sont omises pour des raisons de collinéarité. En effet, ces variables ne changent pas dans le temps et la méthode avec estimateurs à effets fixes n'est pas en mesure de générer leur coefficient. L'écart-type estimée de l'effet fixe est de 39,78 et celui du terme d'erreur est de 44,80. Ces statistiques seront abordées plus en détail lors de la comparaison entre le modèle de régression linéaire et celui à effets

fixes. Le R-Carré pertinent pour cette régression est le R-Carré « intrinsèque » (within) et il est de 0,4577. De plus, la statistique F générée par Stata immédiatement après la régression teste l'hypothèse nulle que l'effet fixe est égal à zéro. On rejette l'hypothèse nulle. Lorsqu'on teste comme hypothèse nulle si le coefficient la variable explicative principale est nul, soit *sportif*, les résultats du test, présentés dans le Tableau A.3a de l'appendice, démontrent qu'on ne rejette pas l'hypothèse nulle. Déjà, le modèle à effets fixes semble produire des résultats qui se contredisent.

Tableau 3.2 Estimateurs : Régression à effet fixes

XTREG, FE - Étudiants sportifs			
Variables	B estimé	Err, robust	P-value
sportif	-0.87	4.4	0.84
âge	46.09	1.5	0.00
filles	0.00	(omitted)	
revenu élevé	-24.26	11.1	0.03
revenu moyen-élevé	-10.72	8.4	0.20
revenu moyen-faible	-10.17	6.5	0.12
études collégiales	7.38	7.2	0.31
études universitaires	-5.22	8.0	0.52
femme immigrante	0.00	(omitted)	
jeune mère	-29.25	13.1	0.03
mère d'âge moyen	-27.18	11.2	0.02
anglais	-6.74	16.3	0.68
bilingue	3.55	12.6	0.78
petite région	-0.06	10.8	1.00
région moyenne	-11.07	14.2	0.44
grande région	-7.32	12.4	0.55
Ontario	-6.82	39.6	0.86
Les provinces des Prairies	2.63	40.1	0.95
Les provinces de l'Atlantique	-6.82	37.5	0.86
Colombie-Britannique	13.24	55.3	0.81
Diplôme Secondaire	23.76	15.0	0.11
Études post-secondaires	18.71	15.5	0.23
Diplôme collégial	12.95	14.7	0.38
Diplôme universitaire	69.04	21.7	0.00
Diplôme collégial	6.99	8.9	0.43
Diplôme universitaire	28.44	21.8	0.19
1 frère/sœur	12.18	9.5	0.20
2 frères/sœurs	16.02	13.6	0.24
3 frères/sœurs	34.12	23.6	0.15
école privée	27.13	19.2	0.16
école catholique	-1.79	8.7	0.84
autres types d'écoles	-13.47	14.2	0.34
famille monoparentale	-12.22	12.6	0.33
2 parents biologiques	1.60	14.9	0.92
constante estimée	-2.21	50.0	0.97

Homme{

N= 1296
Nombre de groupes = 2
Observations par groupes = 2
R^2 within = 0.7637
R^2 between = 0.0492
R^2 overall = 0.4577
σ effet aléatoire = 44.80
σ terme d'erreur = 39.78
Rho (fraction de la variance due à u_i) = 0.5591

Source : ELNEJ et calculs de l'auteur.

L'estimateur de l'impact du sport sur les résultats en math, soit la variable sportif, a un impact négatif de 0,87 point sur le score et n'est pas significatif. L'âge de l'enfant a un impact significatif positif de 46,09 points. Ce résultat est cohérent avec celui généré par le modèle à effets aléatoires. Pour les variables de revenu, les résultats sont incohérents avec la littérature. Pour toutes les variables dichotomiques du revenu familial, l'impact sur le score est négatif par rapport à un étudiant dans un foyer à bas revenu. Le niveau d'étude que les parents espèrent voir leur enfant atteindre n'est pas significatif pour toutes les variables dichotomiques de ce modèle. Tout de même, l'estimateur qui représente des études collégiales complétées augmente la note de 7,38 points par rapport à un qui compléterait des études primaires ou secondaires. Pour des études universitaires, le coefficient est négatif et diminue la note de 5,22 points. Les coefficients des variables dichotomiques relatives à l'âge de la mère de l'étudiant telles que *jeuneMere* et *avgMere* sont négatifs aussi. Pour l'échantillon de la régression, le score diminue de 29,25 points et 27,18 points respectivement par rapport à un élève dont la mère est âgée de 45 ans ou plus. Ces estimateurs sont significatifs à 2 % respectivement. L'estimateur de la variable dichotomique représentant les étudiants anglophones n'est pas significatif et ce constat tient pour celui de la variable représentant les enfants bilingues. Sommes toutes, la note d'un étudiant anglophone accuse 6,74 points de retard sur celle d'un étudiant francophone ou allophone, alors que le coefficient de la variable *biling* estime une augmentation de la note de 3,55 points. Ces résultats sont semblables à ceux obtenus avec le modèle précédent.

En ce qui a trait aux estimateurs des variables dichotomiques qui captent la taille de la région dans laquelle l'enfant réside, ceux qui habitent une zone rurale ont de meilleures notes au test de mathématique que les autres jeunes et ce, que ces derniers habitent une petite région, une moyenne ou une grande. En effet, la note diminue de 0,06 points pour les enfants en petite région et de 11,07 et 7,32 points respectivement

pour ceux qui vivent dans des régions moyennes et grandes. Dans cette régression à effets fixes, l'impact d'habiter dans une province autre que le Québec n'est pas aussi important que dans le modèle à effets aléatoires. Par rapport au score des étudiants québécois, ce dernier diminue de 6,82 points pour les étudiants ontariens et des provinces de l'Atlantique, mais il augmente de 2,63 points pour ceux des Prairies et 13,24 points pour les étudiants qui résident dans la province la plus à l'ouest. Tous ces derniers coefficients ne sont pas significatifs.

En ce qui a trait à l'éducation des parents, contrairement au modèle de l'équation (1A), les coefficients ne sont pas significatifs, sauf dans le cas d'une femme avec un diplôme universitaire. Il s'agit d'ailleurs de la variable de cette catégorie qui améliore le plus la note de l'enfant, soit de 69,04 points versus la note d'un enfant dont la mère n'a pas répondu à la question, c'est-à-dire, lorsque cette dernière variable est égale à zéro. Le résultat augmente de 23,75 points pour les enfants dont la mère possède un diplôme d'études secondaires par rapport à ceux dont la femme est non-répondante. Une variation positive est aussi notée lorsque la mère a complété des études post-secondaires, soit 18,71 points et 12,95 pour celles ayant un diplôme d'études collégiales. Pour l'homme dans le foyer, le résultat augmente de 6,99 points pour un étudiant dont le père a un diplôme collégial par rapport à un enfant dont l'homme a un diplôme secondaire ou moins. Lorsque le père possède un diplôme universitaire, cette augmentation saute à 28,44 points. En ce qui a trait au type d'établissement scolaire, l'effet de suivre un parcours privé versus un public augmente la note de 27,13 points. Cependant, les étudiants des établissements catholiques et autres ont des résultats inférieurs à ceux qui fréquentent l'école publique, soit 1,79 points et 13,47 points respectivement.

Finalement, pour la composition du ménage de l'enfant, les estimateurs des variables dichotomiques générées pour capter l'effet de la taille de la fratrie ne sont pas significatifs. L'ampleur de l'impact de ces variables est plus important dans ce

modèle que dans celui de la section précédente. La note augmente de 12,18 points si l'enfant a un frère ou une sœur à comparer l'enfant unique, 16,02 points s'il y a 3 enfants au total dans le foyer et 34,12 points s'il y a 4 enfants ou plus dans le ménage. Pour un enfant avec un seul parent, le coefficient de cette variable dichotomique propose une baisse de 12,22 points par rapport à un enfant qui a deux parents. Dans cette régression, ce coefficient n'est pas significatif, mais le sens de la variation reste cohérent. Pour terminer, le résultat au test de math augmente de 1,60 points pour un enfant vivant avec ses parents biologiques versus un jeune dont un des deux parents dans son foyer ne l'est pas.

3.1.3 Résultats : Test Hausman

L'hypothèse nulle du test de Hausman préconise le modèle à effets aléatoires, tandis que l'hypothèse alternative favorise le modèle à effets fixes. Ce test se base sur l'hypothèse que l'effet inobservé, c_i , ne peut être corrélé avec le terme d'erreur. Cette hypothèse n'est requise que pour le modèle à effet aléatoire. Ainsi, si cette condition est respectée, les estimateurs des deux modèles seront sans biais, mais le modèle à effet aléatoire sera retenu car il génère une variance plus faible pour les estimateurs. Dans le cas contraire, la régression à effet fixe est le meilleur modèle. Les résultats du test de Hausman sont présentés dans le Tableau A. 4 en appendice. Les régressions sont effectuées sans correction pour l'hétéroscédasticité pour les fins du test. Les coefficients des estimateurs pour le modèle à effets aléatoires pour le test sont donc légèrement différents que ceux générés avec la correction. L'hypothèse nulle est l'absence de corrélation entre l'effet aléatoire et les variables explicatives pour tous les individus à toutes les périodes. Pour cette régression, l'hypothèse nulle n'est pas rejetée. Ceci insinue que les modèles sont potentiellement sans biais, mais étant

donné que la méthode à effets aléatoires estime une variance plus faible pour les estimateurs, cette dernière est donc favorisée pour ce mémoire.

Un autre test, supérieur au test de Hausman, qui aurait pu servir à la validation des modèles est le test de Wald conjoint proposé par Wooldridge à l'aide des moyennes temporelles des variables explicatives (Wooldridge, 2002). Cependant, ce test n'a pas été retenu en raison des limites de temps.

3.1.4 Résultats : Modèle de régression linéaire

Les résultats du modèle de régression sont comparables à ceux obtenus pour le modèle de régression à effets aléatoires pour la majorité des coefficients. Le R-Carré pour chacun des modèles aussi est analogue. Quant à l'estimateur de la variable de sport, tel que discuté plus tôt, l'impact estimé de la pratique d'activité physique sur le score à l'examen de mathématique est de 0,91 point pour le modèle de l'équation (1A). Lorsqu'estimé avec le modèle de régression linéaire simple, cet estimé augmente à 1,25 points, ce qui n'est pas une différence notable. Les résultats de cette régression sont présentés dans le Tableau 3.3 ci-dessous. Le test d'autocorrélation à l'aide du résidu généré par la méthode de régression linéaire permet de régresser le terme d'erreur estimé sur ce même terme d'erreur retardé d'une période afin d'infirmer ou de confirmer la présence d'autocorrélation dans le modèle. Le coefficient de l'erreur retardée estimée est de 0,3094 et ceci soutient du même coup la présence d'autocorrélation dans le terme d'erreur.

Tableau 3.3 Estimateurs : Régression linéaire

Régression linéaire - Étudiants sportifs			
Variabes	B estimé	Err, robust	P-value
sportif	1.25	2.95	0.7
âge	48.91	1.48	0
filie	-0.95	2.75	0.7
revnu élevé	7.21	6.44	0.3
revenu moyen-élevé	5.67	4.91	0.2
revenu moyen-faible	-2.14	4.13	0.6
études collégiales	13.44	5.43	0
études universitaires	11.77	5.99	0
femme immigrante	14.06	5.3	0
jeune mère	-0.58	6.85	0.9
mère d'âge moyen	-4.39	6.74	0.5
anglais	-4.75	4.53	0.3
bilingue	-23.66	10.58	0
petite région	0.95	3.89	0.8
région moyenne	-6.58	4.97	0.2
grande région	-10.17	3.75	0
Ontario	-24.52	6.36	0
Les provinces des Prairies	-27.42	6.5	0
Les provinces de l'Atlantique	-24.7	6.12	0
Colombie-Britannique	-13.77	7.52	0.1
Diplôme Secondaire	15.83	4.96	0
Études post-secondaires	23.77	4.63	0
Diplôme collégial	15.1	4.99	0
Diplôme universitaire	36.24	6.02	0
Homme { Diplôme collégial	8.24	3.5	0
Diplôme universitaire	2.47	4.67	0.6
1 frère/sœur	4.45	4.72	0.3
2 frères/sœurs	7.83	5.01	0.1
3 frères/sœurs	3.53	6.57	0.6
école privée	24.89	8.99	0
école catholique	7.98	3.9	0
autres types d'écoles	5.46	21.74	0.8
famille monoparentale	-3.34	7.39	0.7
2 parents biologiques	1.66	6.29	0.8
constante estimée	-37.81	17.36	0

N= 1296
R ² = 0.5480
R ² ajusté = 0.5358

Source : ELNEJ et calculs de l'auteur.

3.2 Résultats : Modèles modifiés

3.2.1 Résultats : Modèles à effets aléatoires

Lorsque l'équation (1A) est régressée, mais en imposant *actif* comme variable explicative principale, on obtient l'équation (1A').

Soit l'équation (1A') :

$$y_{it} = \beta_0 + \beta_1 Z_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} + c_i + u_{it}, \quad (1A')$$

où y_{it} est le score au test de mathématique, Z_{1t} est la variable *actif*, x_{it} sont les toutes variables explicatives retenues jusqu'à présent, c_i est l'effet aléatoire, u_{it} est le terme d'erreur, i est l'identifiant de l'enfant et t est l'indice de temps.

Les coefficients des variables de contrôle estimés par ces régressions sont très semblables, c'est-à-dire que l'ampleur et le sens des variations sont pratiquement les mêmes. Dans les cas les plus extrêmes, les coefficients des deux modèles diffèrent au maximum par 0,2 points. Par contre, la constante et l'estimateur de la variable *actif* propose un impact plus important, soit de 3,15 points pour la variable explicative principale. Pourtant, la corrélation entre la variable *scoreMath* et *sportif* et plus grande que celle entre *scoreMath* et *actif*. Il est à noter que les estimateurs qui n'étaient pas significatifs dans le modèle pour les étudiants sportifs ne le sont pas dans ce modèle-ci aussi. Le R-Carré et les écart-type estimés sont aussi semblables, d'où des résultats similaires lorsque le test de Breush-Pagan est conduit sur les deux modèles. Le résultat du test, présenté dans le Tableau A.2 en appendice, permet de rejeter l'hypothèse nulle. Les résultats de la régression de l'équation (1A') eux sont présentés dans le Tableau 3.4 ci-dessous.

Tableau 3.4 Estimateurs : Régression à effets aléatoires – actifs

XTREG, RE- Étudiants actifs			
Variables	B estimé	Err, robust	P-value
sportif	3.15	3.2	0.87
âge	48.19	1.2	0
filie	-1.06	3.1	0.73
revenu élevé	2.42	7	0.73
revenu moyen-élevé	2.96	4.9	0.55
revenu moyen-faible	-3.75	4.2	0.38
études collégiales	12.54	5.6	0.03
études universitaires	7.23	6.2	0.25
femme immigrante	13.73	5.6	0.01
jeune mère	-4.13	8.5	0.63
mère d'âge moyen	-7.28	8.5	0.39
anglais	-4.82	5.4	0.37
bilingue	-18.59	9.2	0.04
petite région	1.32	4.5	0.77
région moyenne	-6.65	5.3	0.21
grande région	-9.53	4.2	0.02
Ontario	-24.26	6.7	0
Les provinces des Prairies	-27.16	7.4	0
Les provinces de l'Atlantique	-25	6.4	0
Colombie-Britannique	-13.47	7.9	0.09
Diplôme Secondaire	17.22	5.4	0
Études post-secondaires	24.46	5	0
Diplôme collégial	16.15	5.4	0
Diplôme universitaire	38.26	6.5	0
Homme { Diplôme collégial	8.59	3.6	0.02
Homme { Diplôme universitaire	3.02	5.2	0.56
1 frère/sœur	5.16	5.4	0.34
2 frères/sœurs	8.24	5.6	0.14
3 frères/sœurs	4.54	7.6	0.55
école privée	23.64	9.5	0.01
école catholique	7.27	4.1	0.08
autres types d'écoles	1.03	6.6	0.88
famille monoparentale	-4.32	7.8	0.58
2 parents biologiques	2.19	6.9	0.75
constante estimée	-27.27	17.3	0.12

N= 1296
Nombre de groupes = 2
Observations par groupes = 2
R ² within = 0.7552
R ² between = 0.1674
R ² overall = 0.5471
σ effet aléatoire = 26.50
σ terme d'erreur = 39.70
Rho (fraction de la variance due à u _i) = 0.3082

Source : ELNEJ et calculs de l'auteure.

3.2.2 Résultats : Modèle à effets fixes

Si l'équation (1B) est régressée avec *actif* comme variable explicative principale, elle devient l'équation (1B').

Soit l'équation (1B') :

$$y_{it} = \beta_0 + \beta_1 Z_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} + D_1 c_1 + D_2 c_2 + \dots + D_n c_n + u_{it}, \quad (1B')$$

où y_{it} est le score au test de mathématique, Z_{1t} est la variable *actif*, x_{it} sont les variables explicatives, c_i est une variable non-observée qui ne varie pas dans le temps, soit l'effet fixe, D_i est une des variables dichotomiques qui prend la valeur de 1 ou 0, u_{it} est un terme d'erreur, i est l'identifiant de l'enfant et t est l'indice de temps.

Les coefficients des variables de contrôle estimés par cette régression modifiée, mise à part celui de la constante et celui de la variable *actif*, sont d'ampleur similaire à ceux de l'équation (1B) et le sens des variations sont les mêmes pour la majorité des variables explicatives pour le deux modèles. Ces résultats sont présentés dans le Tableau 3.5 ci-dessous. Parmi les valeurs qui sont différentes dans ce modèle, l'effet du fait d'habiter une petite région au lieu d'une région rurale est positif alors qu'il était négatif pour le modèle de régression estimé à l'aide de l'équation (1B). Un constat similaire s'applique à la constante qui passe de -2,21 points dans le modèle de base à 7,66 points pour l'équation modifiée. Aussi, le coefficient de la variable explicative principale passe de -0,87 point avec *sportif* à -6,65 points avec *actif*. Ici, l'impact d'être inscrit à une activité scolaire quelconque fait varier la note de façon négative par rapport à ceux qui ne le sont pas, et cette variation est plus importante que dans le modèle pour les sportifs. Finalement, le coefficient de la variable *FemmeDiplSec*, soit une mère ayant obtenu un diplôme d'études secondaires, est significatifs à 9 % tandis que ce pourcentage était de 11 % dans le modèle de base. Encore une fois, les variables dichotomiques *ageEnfant* et *immigrantFemme* ont été omises pour éviter le problème de collinéarité. La statistique F générée après la régression pour tester

l'hypothèse nulle que tous les effets fixes sont égaux à zéro nous permet de rejeter cette dernière hypothèse nulle. Lorsqu'on teste comme hypothèse nulle si le coefficient de la variable explicative principale est nul, soit *actif*, les résultats du test, présentés dans le Tableau A.3b de l'appendice, démontrent qu'on ne rejette pas l'hypothèse nulle à un niveau de significativité de 5 %, mais il faut noter qu'à 11 %, on la rejette.

Tableau 3.5 Estimateurs : Régression à effets fixes – actifs

XTREG, FE- Étudiants actifs			
Variables	B estimé	Err, robust	P-value
sportif	-6.65	4.1	0.1
âge	46.04	1.4	0
filie	0	(omitted)	
revenu élevé	-25.09	11	0.02
revenu moyen-élevé	-10.93	8.3	0.19
revenu moyen-faible	-10.27	6.4	0.11
études collégiales	7.92	7.3	0.28
études universitaires	-4.51	8	0.57
femme immigrante	0	(omitted)	
jeune mère	-30.54	13	0.02
mère d'âge moyen	-28.21	11.2	0.01
anglais	-7.87	16.5	0.63
bilingue	2.46	12.9	0.85
petite région	0.17	10.8	0.99
région moyenne	-9.88	14.4	0.49
grande région	-6.61	12.5	0.6
Ontario	-8.69	37.9	0.82
Les provinces des Prairies	1.93	38.2	0.96
Les provinces de l'Atlantique	-10.37	35.4	0.77
Colombie-Britannique	12.25	53.8	0.82
Diplôme Secondaire	25.41	14.8	0.09
Études post-secondaires	17.75	15	0.24
Diplôme collégial	11.96	14.4	0.41
Diplôme universitaire	66.46	22.1	0
Diplôme collégial	6.34	9	0.48
Diplôme universitaire	26.68	22.6	0.24
1 frère/sœur	11.27	9.3	0.22
2 frères/sœurs	14.92	13.4	0.27
3 frères/sœurs	32.28	24.1	0.18
école privée	23.85	20.1	0.24
école catholique	-1.93	8.8	0.83
autres types d'écoles	-12.15	14.1	0.39
famille monoparentale	-14.32	12.5	0.25
2 parents biologiques	1.7	14.6	0.91
constante estimée	7.66	48.9	0.88

Homme {

N= 1296
Nombre de groupes = 2
Observations par groupes = 2
R^2 within = 0.7647
R^2 between = 0.0470
R^2 overall = 0.4596
σ effet aléatoire = 44.65
σ terme d'erreur = 39.70
Rho (fraction de la variance due à u_i) = 0.5584

Source : ELNEJ et calculs de l'auteur.

3.2.3 Résultats : Test Hausman

Comme pour les régressions avec $Z_{it} = sportif$, le même test de Hausman est utilisé pour sélectionner le modèle le plus adéquat pour les données disponibles. L'hypothèse nulle du test de Hausman préconise le modèle à effet aléatoire, tandis que l'hypothèse alternative favorise le modèle à effet fixe. Les résultats du test de Hausman sont présentés dans le Tableau A. 5 en appendice. Les régressions sont effectuées sans correction pour l'hétéroscédasticité pour les fins du test. Les coefficients des estimateurs pour le modèle à effet aléatoire sont donc légèrement différents que ceux générés avec la correction. Pour cette régression et pour un niveau de significativité de 5 %, l'hypothèse nulle n'est pas rejetée. Ceci implique que les modèles sont potentiellement sans biais, mais pour ce mémoire, le modèle à effets aléatoires est préféré pour la variance estimée de ses estimateurs plus faible.

3.2.4 Résultats : Modèle de régression linéaire

Quand l'équation (1C) est régressée avec *actif* comme variable explicative principale, elle devient l'équation (1C').

Soit l'équation (1C) :

$$y_i = \beta_0 + \beta_1 Z_1 + \beta_2 x_2 + \dots + \beta_n x_n + u_i, \quad (1C)$$

qui consiste en l'équation (1), mais sans l'indice de temps et où y_{it} est le score au test de mathématique, Z_{1t} est la variable *actif*, x_{it} sont les variables explicatives, et u_{it} est un terme d'erreur.

Les résultats du modèle de l'équation (1C') sont comparables, pour la majorité des coefficients, à ceux obtenus pour les modèles de régression des équations (1A), (1A')

et (1C). Le R-Carré aussi est équivalent pour ces modèles. La variation positive estimée de la note, pour un jeune qui pratique une activité parascolaire quelconque, est de 2,59 points par rapport à un étudiant inscrit à aucune activité. Les résultats de cette régression sont présentés dans le Tableau 3.6 ci-après.

Le test d'autocorrélation du modèle linéaire cherche à estimer un coefficient pour l'erreur estimée par ce même modèle sur sa valeur à la période précédente. Un estimateur non nul confirme la présence d'autocorrélation dans le modèle, ce qui est le cas pour le modèle généré avec l'équation (1C'). Ce dernier est estimé à 0,3065 et confirme que le terme d'erreur est corrélé avec sa valeur passée.

Tableau 3.6 Estimateurs : Régression linéaire - actifs

Régression linéaire - Étudiants actifs			
Variables	B estimé	Err, robust	P-value
sportif	2.59	3.12	0.4
âge	48.91	1.48	0
filie	-0.97	2.71	0.7
revenu élevé	7.4	6.4	0.2
revenu moyen-élevé	5.84	4.86	0.2
revenu moyen-faible	-2.11	4.1	0.6
études collégiales	13.32	5.43	0
études universitaires	11.45	6	0.1
femme immigrante	14.22	5.3	0
jeune mère	-0.59	6.85	0.9
mère d'âge moyen	-4.34	6.74	0.5
anglais	-4.84	4.52	0.3
bilingue	-23.79	10.58	0
petite région	0.92	3.89	0.8
région moyenne	-6.78	4.98	0.2
grande région	-10.25	3.75	0
Ontario	-24.26	6.36	0
Les provinces des Prairies	-27.2	6.48	0
Les provinces de l'Atlantique	-24.73	6.12	0
Colombie-Britannique	-13.69	7.51	0.1
Diplôme Secondaire	15.81	4.95	0
Études post-secondaires	23.8	4.62	0
Diplôme collégial	15.08	4.97	0
Diplôme universitaire	36.24	6	0
Homme { Diplôme collégial	8.33	3.48	0
Diplôme universitaire	2.35	4.66	0.6
1 frère/sœur	4.61	4.72	0.3
2 frères/sœurs	7.92	5.01	0.1
3 frères/sœurs	3.61	6.57	0.6
école privée	25.16	8.99	0
école catholique	8.02	3.9	0
autres types d'écoles	4.95	21.75	0.8
famille monoparentale	-3.02	7.37	0.7
2 parents biologiques	1.81	6.28	0.8
constante estimée	-39.18	17.42	0

N= 1296
R ² = 0.5482
R ² ajusté = 0.5360

Source : ELNEJ et calculs de l'auteur.

CHAPITRE IV

DISCUSSION DES RÉSULTATS

Dans cet essai qui tente d'estimer l'impact du sport sur les résultats au test de mathématique, « l'étudiant-type » pour cet échantillon, c'est-à-dire lorsque toutes les variables dichotomiques sont égales à zéro, est un garçon dans un ménage où le revenu familial est bas, soit sous les 30 000 \$, pour lequel les parents croient qu'il complètera ses études secondaires et qui a une mère âgée de 45 ans ou plus qui est née au Canada. Le garçon est francophone ou allophone, mais pas anglophone ni bilingue. Il habite dans une région rurale du Québec et est enfant unique. Le père a complété des études secondaires ou moins alors que le niveau maximum d'études atteint par la mère n'a pas été déclaré. D'ailleurs, il grandit dans un foyer avec deux parents, mais seul un des deux est son parent biologique. Les tableaux B.1 et B.2 de l'appendice présentent les statistiques descriptives des variables utilisées dans les régressions pour les étudiants sportifs et actifs respectivement.

Dans le modèle de régression avec estimateurs à effets aléatoires, parmi les variables explicatives incluses dans l'équation (1A), celles qui influencent le plus le score, toutes choses étant égales par ailleurs, sont si la mère est née au Canada ou non, le niveau d'éducation de cette dernière et le type d'établissement que fréquente l'enfant. Pour cette méthode, le résultat moyen d'un étudiant qui va à l'école privé augmente de 6,51 % à comparer un enfant qui fréquente une école publique. Ce constat est cohérent avec la littérature. Pour ce qui a trait à la variable relative au statut d'immigrant de la mère, le coefficient estime une augmentation de 3,77 % de la note moyenne pour un étudiant dont la mère n'est pas née au Canada par rapport à un jeune pour qui la mère l'est. Toujours au sujet de la mère, il a été estimé que celles

qui ont obtenu leur diplôme universitaire font augmenter la note de l'enfant de 10,47 % versus une mère non-répondante. Pour cet échantillon, les caractéristiques relatives à la mère influencent beaucoup la note de l'enfant lorsque les effets inobservés sont considérés aléatoires.

Dans ce même modèle de régression, toutes les variables représentant le revenu du foyer ont une influence positive sur la note par rapport à un étudiant vivant dans un ménage à bas revenu sauf dans le cas où le revenu est considéré « moyen-bas ». Une hypothèse qui pourrait être soulevée pour expliquer l'estimateur négatif généré pour cette variable est que le fait que les foyers dans cette catégorie ont un revenu trop élevé pour bénéficier d'assistance financière pour, par exemple, les frais de scolarité, le tutorat et les activités parascolaires, mais trop faible pour financer ces besoins eux-mêmes. Il s'ensuit que la note moyenne d'un jeune qui vit dans ce genre de ménage diminue de 1,05 % par rapport à un étudiant dont le revenu familial est sous le seuil des 30 000 \$. Ce genre de dilemme se pose aussi dans le financement sportif. Que ce soit pour une bourse Sport-Études ou une bourse pour les jeunes espoirs, le revenu familial constitue un critère de sélection important pour octroyer ces sommes. Ceci permet de favoriser les plus démunies certes, mais comme mentionné plus haut, les familles qui sont dans l'impossibilité de faire l'investissement initial sans soutien financier ne peuvent pas inscrire leur enfant dans des activités hors-école. Par exemple, la fondation Bon Départ de Canadian Tire affirme sur son site web qu'elle a pour objectif fondamental de venir en aide aux enfants de familles financièrement désavantagées afin qu'ils puissent eux aussi prendre part à des sports ou des activités récréatives. Si à la source les programmes sportifs coûtent moins chers, soit en subventionnant les écoles qui mettent en place des équipes sportives ou les municipalités qui offrent des ligues, le sport devient plus accessible aux jeunes.

Dans ce mémoire, l'impact de la pratique d'une activité physique sur le résultat au test de mathématique, par rapport à la note moyenne obtenue pour les étudiants de l'échantillon est négligeable dans la régression à effets aléatoires, soit une augmentation de 0,25 % du score. Le R-Carré « overall » qui est une mesure de la capacité du modèle à expliquer les variations dans les notes au test est de 0,5471. Ainsi, le modèle de l'équation (1A) permet d'expliquer partiellement la diversité des résultats. Plusieurs explications sont possibles. D'abord, pour les enfants de moins de 10 ans, ce sont les parents qui répondent aux questions de l'ELNEJ pour eux. Ensuite, les individus observés sont peut-être encore trop jeunes pour pratiquer un sport à un niveau assez élevé qui leur permettrait de bénéficier de ces bienfaits possibles. Finalement, dans l'étude de Trois-Rivière (Shepard, 1997), les auteurs mentionnent qu'en première année, les jeunes dans le groupe avec le nombre minimum d'heures d'éducation physique obtenaient de meilleurs résultats que ceux dans le groupe avec une heure supplémentaire. Cependant, cette tendance était renversée en fin d'études primaires.

Dans le modèle de régression avec estimateurs à effets fixes estimé avec l'équation (1B), plusieurs coefficients estiment des variations incohérentes avec la littérature. D'abord, tous les estimateurs des variables de revenu sont négatifs et celui pour la variable *sportif* aussi. De plus, le coefficient de cette dernière variable estime une diminution de la note de 0,239 %. Le R-Carré « within » pour cette régression est de 0,7637, ce qui à première vue porte à croire que cette modélisation permet de bien expliquer les variations du score de l'étudiant. Néanmoins, le test de Hausman ne permet pas de rejeter l'hypothèse nulle, mais il favorise tout de même le modèle à effets aléatoires. Plusieurs raisons justifient cette conclusion, mais la plus importante est que les variables incluses dans l'équation (1B) ne varient pas assez dans le temps pour que leur impact sur la note soit bien capté par le modèle à effets fixes dans le test de Hausman (Merrigan, 2014). D'ailleurs, les variables *sexeEnfant* et *immigrantFemme* sont omises dans l'estimation de cette équation.

Lorsque la variable explicative principale est remplacée par *actif* dans le modèle de régression à effets aléatoires, on remarque que les estimateurs de plusieurs variables de l'équation (1A') sont, au dixième près, identiques à ceux de l'équation (1A) sauf pour le coefficient de la variable *actif*. Ce dernier estime une variation positive de 0,86 % du score pour un étudiant actif par rapport à un qui ne l'est pas. Malgré une corrélation faible entre les deux variables explicatives principales, leur impact respectif sur la note est de moins de 1 % dans les deux cas. Quand même, la catégorie pour les enfants actifs est beaucoup plus vaste que pour les sportifs, ce qui fait qu'une variation un peu plus importante de la note est attendue et ce, même si la corrélation entre le score et le sport est plus importante que celle entre la note et une activité quelconque. Le R-Carré « overall » est le même dans l'équation (1A) et (1A'), mais les écart-types pour le terme d'erreur et l'effet aléatoire sont plus petites dans le modèle modifié. Quand même, le test de Breush-Pagan effectué sur le modèle de l'équation (1A') génère les mêmes conclusions, c'est-à-dire qu'on rejette une variance nulle pour l'effet aléatoire.

Les comparaisons sont similaires entre l'équation (1B) et (1B') qui sont estimées avec une régression à effets fixes. Dans le modèle avec la variable *actif*, le coefficient de cette dernière variable estime une variation négative de 1,82 % de la note pour un enfant qui pratique une activité parascolaire quelconque par rapport à un pour qui ce n'est pas le cas. Le R-Carré pour le modèle modifié est de 0,7647, ce qui est semblable au modèle de base. Les écart-type estimés par l'équation (1B') sont plus faibles que ceux générés par la méthode initiale. Ainsi, lorsque Stata teste l'hypothèse nulle que tous les effets fixes sont égaux à zéro, la statistique F générée est sensiblement différente, mais on rejette tout de même l'hypothèse nulle pour l'équation (1B'). Le test sur le coefficient de la variable explicative principale démontre qu'on ne rejette pas l'hypothèse nulle que cet estimateur est nul. Cependant, dans le cas de la variable *actif*, à 11 %, on rejette l'hypothèse nulle.

D'ailleurs, parmi les modèles présentés précédemment, seuls ceux avec des estimateurs à effets fixes calculent que l'impact sur la note au test de mathématique de pratiquer une activité, sportive ou non, hors de l'établissement scolaire est négatif. Cette méthode calcule aussi, en terme absolu, des variations plus importantes du score. Ces observations sont cohérentes avec la littérature, dont l'étude de Sallis *et al.* (Sallis, 1997) présenté dans la section Introduction. Au cours d'une journée, le temps accordé à la pratique d'activités parascolaires diminue celui destiné aux études et aux devoirs, toutes choses étant égales par ailleurs. Ainsi, il est possible, surtout en bas âge, qu'une mauvaise gestion des heures attribuées à chaque tâche pénalise l'enfant dans son apprentissage, d'où une variation négative dans la note pour les étudiants sportifs ou actifs.

À la lumière de ces résultats, le modèle de régression avec estimateurs à effets aléatoires sera retenu pour ce mémoire. Plusieurs raisons, dont le test de Hausman qui favorise le modèle à effets aléatoires et les résultats négatifs proposés par le modèle à effets fixes justifient ce choix.

Pour comparer les trois modèles présentés jusqu'ici dans cette recherche, il est possible de calculer un ratio qui les unit dans les cas limites, comme mentionné dans la présentation des modèles.

Soit l'équation (2) :

$$\theta = 1 - \sqrt{\sigma_u^2 / (T\sigma_e^2 + \sigma_u^2)}, \quad (2)$$

où θ est la proportion de la variance totale qui provient de l'effet inobservé, aussi appelé teta dans le modèle de régression estimé avec Stata, σ_u^2 est la variance du terme d'erreur, T est l'indice de temps et σ_e^2 est la variance de l'effet inobservé. Ces

résultats pour le modèle de régression à effets aléatoires pour les sportifs sont présentés dans le Tableau 4.1 suivant.

Tableau 4.1 Données équation (2)

RE, Sportif
Teta = 0,3080
sigma (c) = 26,54
sigma (u)= 39,78

Source : ELNEJ et calculs de l'auteur.

Le cas limite du modèle de régression à effets aléatoires lorsque θ est égal à zéro, est le modèle de régression linéaire estimé par l'équation (1C). À l'opposé, si θ est égal à un, on obtient le modèle de régression à effets fixes (Merrigan, 2014). Pour le modèle retenu, $\theta = 0,2726$ et $teta = 0,3080$. Le modèle de régression avec estimateurs à effets aléatoires est donc adéquat. Ceci explique sans doute pourquoi les coefficients de l'équation (1C) sont généralement plus élevés que ceux de l'équation (1A). En effet, si la méthode par moindre carré ordinaire ne tient pas compte de la corrélation intertemporelle entre les variables explicatives et les effets inobservés, elle génère des estimateurs biaisés.

CONCLUSION

Le but de cette recherche était de déterminer s'il y avait un lien de causalité entre l'activité physique et les résultats académiques et si ce lien était significatif. Cet objectif a été partiellement atteint. Avec un panel tiré de l'Enquête longitudinale nationale sur les enfants et les jeunes qui couvre les années 1994-1995 à 2008-2009, pour tous les étudiants canadiens, il a été possible d'extraire un échantillon de 1300 étudiants âgés de 7 et 9 ans pour les trois premiers cycles de cette enquête. Malgré les problèmes d'hétéroscédasticité et d'autocorrélation qui sont faibles, mais qui perdurent, une relation positive mais non significative a été estimée par le modèle de régression à effets aléatoires.

La conclusion est que l'impact de l'activité physique sur les résultats scolaires, plus précisément, sur la note de mathématiques moyenne, est de 0,25 %. Plusieurs raisons expliquent cet impact faible, mais notable. D'abord, les hypothèses strictes du modèle à effets aléatoires amènent avec elles leurs lots de complications comme l'hétéroscédasticité ou l'autocorrélation du terme d'erreur. Ainsi, la nature de l'effet inobservé et sa modélisation, qui sont eux-mêmes des sujets de recherches importants, influencent la qualité des résultats qui peuvent être tirés de ce genre de régression. Ensuite, les variables relatives au temps dédié aux sports au lieu des devoirs, les heures de sommeil et le passé sportif des parents sont des variables qui sont absentes du modèle, car elles ne faisaient pas parties de l'enquête, mais qui pourtant ont sans doute un impact sur les performances autant sportives que académiques. Puis, la littérature démontre que les jeunes sportifs qui performant mieux académiquement sont plus âgés que ceux retenus dans ce mémoire. De plus, le fait que la même question ne soit pas répétée pour les adolescents limite la taille de l'échantillon. Finalement, comme mentionné au Chapitre 1, le faible taux de réponse

à l'ELNEJ diminue l'information disponible pour les chercheurs par le fait que les caractéristiques des enfants non-répondants ne sont pas captées dans la régressions et cela pourrait causer un biais dans les estimations.

Par conséquent, la motivation la plus importante de ce travail, soit le financement du développement humain par l'intermédiaire du sport est donc partiellement validée. Non seulement le sport affecte positivement la note moyenne des étudiants dans cet échantillon, mais l'éducation des parents ainsi que les choix que font ces derniers quant aux études de leur enfant et à la structure de leur foyer influence aussi à la hausse les résultats académiques.

En 1980, le premier programme sport-études est implanté dans la ville de Boucherville au Québec. Au départ, le but premier du programme sport-études n'était pas de décourager le décrochage scolaire, pourtant ce résultat était sans équivoque le plus flagrant. Plusieurs professeurs qui ont pris part au projet de départ affirme maintenant que le succès dans le sport et dans les études vont de paires (ici.radio-canada.ca, 2015).

APPENDICE A

TESTS DE VALIDATION DES MODÈLES DE RÉGRESSION

Tableau A.1 Breush-Pagan : XTREG, RE – Sportifs

Breusch and Pagan Lagrangian multiplier test for random effects

$$\text{scoreMath}[id,t] = Xb + u[id] + e[id,t]$$

Estimated results:

	Var	sd = sqrt(Var)
scoreMath	4932.755	70.23358
e	1582.492	39.78055
u	704.527	26.54293

Test: Var(u) = 0

$$\begin{aligned} \text{chibar2}(01) &= 54.29 \\ \text{Prob} > \text{chibar2} &= 0.0000 \end{aligned}$$

Source : ELNEJ et calculs de l'auteure.

Tableau A.2 Breush-Pagan : XTREG, RE – Actifs

Breusch and Pagan Lagrangian multiplier test for random effects

$$\text{scoreMath}[id,t] = Xb + u[id] + e[id,t]$$

Estimated results:

	Var	sd = sqrt(Var)
scoreMath	4932.755	70.23358
e	1576.152	39.70078
u	702.2741	26.50046

Test: Var(u) = 0

$$\begin{aligned} \text{chibar2}(01) &= 53.41 \\ \text{Prob} > \text{chibar2} &= 0.0000 \end{aligned}$$

Source : ELNEJ et calculs de l'auteur.

Tableau A.3 Test variables explicatives principales

(a) Effets fixes - sportifs

(b) Effets fixes - actifs

sportif = 0

$$\begin{aligned} F(1, 647) &= 0.04 \\ \text{Prob} > F &= 0.8416 \end{aligned}$$

actif = 0

$$\begin{aligned} F(1, 647) &= 2.67 \\ \text{Prob} > F &= 0.1027 \end{aligned}$$

Source : ELNEJ et calculs de l'auteur.

Tableau A.4 Hausman – Sportifs

Note: the rank of the differenced variance matrix (31) does not equal the number of coefficients being tested (32); be sure this is what you expect, or there may be problems computing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) .		
sportif	-1.8705549	.905792	-1.776247	1.012915
ageEnfant	46.08963	48.19039	-2.100769	.747875
RevenuFler6	-24.26194	2.26623	-26.52817	9.446834
RevenuMoyen2	-10.71966	2.776908	-13.49657	6.807483
RevenuMoyen1	-10.16616	-3.89933	-6.27215	4.49428
EtudeColl	7.380414	12.51068	-5.130264	5.091609
EtudeUni	-5.221718	7.251764	-12.47348	5.557305
jeuxHere	-29.25004	-4.160076	-25.08997	11.65082
svHere	-27.17905	-7.336947	-19.8422	9.848292
angl	-6.740164	-4.77437	-1.965793	16.90912
biling	3.54797	-19.50556	22.05353	13.33789
petiteRegion	-.0568041	1.265521	-1.322325	8.16959
regionMoyen	-11.07046	-6.630781	-4.439679	12.33703
grandeRegion	-7.321599	-9.544365	2.222765	14.4705
ontario	-6.822055	-24.35315	17.53109	49.75981
prairie	2.628126	-27.30002	29.92814	47.83281
atlantique	-6.822821	-24.99542	18.1726	40.26378
CB	13.23653	-13.56505	26.80158	58.24621
FemmeDiplSec	23.75521	17.14668	6.608536	14.55506
FemmeEt-tSec	18.70793	24.36408	-5.656169	13.6649
FemmeDiplC-1	12.94615	16.03787	-3.091723	25.46399
FemmeDiplUni	69.03558	18.10581	30.92977	13.38886
HommeDiplC-1	6.887572	8.5821	-1.514528	11.89623
HommeDiplUni	28.43674	2.97559	25.46115	10.43817
frere1	12.18225	5.108797	7.079452	14.79426
frere2	16.02166	8.196254	7.825402	16.70099
frere3	34.11981	4.495762	29.62405	26.23689
privé	27.13364	23.70206	3.43158	27.13878
catholique	-1.788862	7.270424	-9.059286	7.897281
autreEcole	-13.47124	1.094316	-14.56555	15.70903
monoparental	-12.21652	-4.53284	-7.683678	10.15051
parentsBio	1.600482	2.049225	-.4487424	25.05848

b = consistent under H_0 and H_a ; obtained from xtreg
 B = inconsistent under H_a , efficient under H_0 ; obtained from xtreg

Test: H_0 : difference in coefficients not systematic

$$\begin{aligned} \text{chi2}(31) &= (b-B)' [(V_b-V_B)^{-1}] (b-B) \\ &= 39.44 \\ \text{Prob>chi2} &= 0.1975 \\ (V_b-V_B \text{ is not positive definite}) \end{aligned}$$

Source : ELNEJ et calculs de l'auteur.

Tableau A.5 Hausman – Actifs

Note: the rank of the differenced variance matrix (11) does not equal the number of coefficients being tested (12); be sure this is what you expect, or there may be problems competing the test. Examine the output of your estimators for anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.

	Coefficients		(b-B) Difference	sqrt((diag(V_b-V_B)) E.H.
	(b) Fixed_Actifs	(B) .		
actif	-6.84473	.800008	-7.348739	2.879661
ageInfant	46.64289	48.15452	-2.152047	.7429384
RevenuEleve	-25.60072	2.431473	-27.96619	9.466541
RevenuMoyen1	-18.9338	2.54557	-23.88747	6.831183
RevenuMoyen2	-18.24934	-1.754284	-6.51466	4.505654
EtudeColl	7.922922	22.54036	-4.627257	5.099739
EtudeUni	-4.513965	7.233633	-11.74954	5.551578
JeuneMere	-18.53731	-4.127342	-26.46997	11.67169
avgMere	-28.20929	-7.281244	-20.92704	9.831653
angl	-7.845569	-4.824054	-3.043415	16.31459
hiling	2.455388	-18.59025	21.04564	12.34409
pat:telregion	.173067	1.337056	-1.344891	8.179479
regionMoyen	-9.880966	-4.653374	-3.227592	12.35115
grandregion	-8.687084	-5.523057	3.924024	14.4601
ontario	-8.693785	-24.25684	15.56309	49.69437
prairie	1.827179	-27.1579	29.08508	47.81674
atlantique	-18.37273	-25.66442	14.61169	40.25193
CB	12.2449	-21.47438	25.72128	58.23882
FemmeD:plSec	25.40916	17.31908	8.19016	14.58673
FemmeEt:1Sec	17.7455	24.44363	-6.717516	12.67076
FemmeD:plC-1	11.95762	16.34874	-4.191716	15.46106
FemmeD:plUni	66.45504	78.24339	20.19169	23.43344
HommeD:plC-1	6.34163	8.551622	-2.252582	11.90685
HommeD:plUni	26.88059	1.026444	23.66014	30.43247
frere1	11.273	5.266684	6.112318	14.75029
frere2	14.91008	8.243158	6.674922	16.89562
frere3	12.17546	4.54318	17.73251	24.26104
privé	29.65352	23.63684	6.217466	27.213
catholique	-1.934365	7.274881	-9.209246	7.87971
autresRelig	-12.1549	3.031773	-13.34022	19.71497
nonparental	-14.32294	-4.327495	-39.88586	18.26467
parentale	1.484587	2.284493	-.899463	25.0272

b = consistent under H0 and H1; obtained from xtreg

B = inconsistent under H1, efficient under H0; obtained from xtreg

Test: H0: difference in coefficients not systematic

$$\chi^2(11) = (b-B)'(V_b-V_B)^{-1}(b-B)$$

$$= 42.36$$

$$\text{Prob} < \chi^2 = 0.0000$$

Source : ELNEJ et calculs de l'auteure.

APPENDICE B

STATISTIQUES DESCRIPTIVES DES VARIABLES UTILISÉES DANS LES RÉGRESSIONS

Tableau B.1 Statistiques descriptives des variables utilisées dans les régressions pour les étudiants sportifs

Variable	Obs	Mean	Std. Dev.
scoreMath	1300	363.8492	70.16009
actif	1300	.74	.438803
ageEnfant	1300	8	1.000385
sexeEnfant	1300	.5153846	.4999556
RevenuRlevé	1300	-.0915385	-.2884843
RevenuMoyen2	1300	.2515385	-.4340643
RevenuMoyen1	1300	-.4761538	-.4996232
EtudeColl	1300	.6853846	-.4645412
EtudeUni	1300	-.2430769	-.4291062
immigrantK-e	1296	-.0740741	-.2619925
jeuneMere	1300	-.5007692	-.5001918
avgMere	1300	-.4515385	-.4978375
angl	1300	-.7838462	-.4117788
hilling	1300	-.0184615	-.134665
petiteRegion	1300	.2376923	-.4258335
regionMoyen	1300	-.1123077	-.3158661
grandeRegion	1300	.3892308	-.4877636
ontario	1300	.2676923	-.4429267
prairie	1300	-.2523077	-.434504
atlantique	1300	-.2869231	-.4524994
CA	1300	-.0915385	-.2884843
FemmeDiplSec	1300	.2015385	-.4013036
FemmeEL-LSec	1300	.3115385	-.4633005
FemmeDiplC-1	1300	-.2261538	-.4185009
FemmeDiplUni	1300	.1307692	-.3372775
HommeDiplC-1	1300	-.2238462	-.4169806
HommeDiplUni	1300	.1353846	-.3422656
frere1	1300	.51	-.5080924
frere2	1300	.3053846	-.4407474
frere3	1300	.0784615	.269
privé	1300	-.0246154	-.1550095
catholique	1300	-.1538462	-.3609401
autrescole	1300	.0038462	-.0619218
monoparental	1300	-.1192308	-.3211846
parentsBiv	1300	-.8292308	-.376452

Source : ELNEJ et calculs de l'auteure.

Tableau B.2 Statistiques descriptives pour les variables utilisées dans les régressions pour les étudiants actifs

Variable	Obs	Mean	Std. Dev.
scoreMath	1300	363.8492	70.16009
actif	1300	.74	.438803
ageEnfant	1300	8	1.000385
sexeEnfant	1300	.5153846	.4999556
RevenuElevé	1300	.0915385	.2884843
RevenuMoyen2	1300	.2515385	.4340643
RevenuMoyen1	1300	.4761538	.4996232
EtudeColl	1300	.6853846	.4645412
EtudeUni	1300	.2430769	.4291062
immigrantP-e	1296	.0740741	.2619925
jeuneMere	1300	.5007692	.5001918
avgMere	1300	.4515385	.4978375
angl	1300	.7838462	.4117788
biling	1300	.0184615	.134665
petiteRegion	1300	.2376923	.4258335
regionMoyen	1300	.1123077	.3158661
grandeRegion	1300	.3892308	.4877635
ontario	1300	.2676923	.4429267
prairie	1300	.2523077	.434504
atlantique	1300	.2869231	.4524994
CB	1300	.0915385	.2884843
FemmeDiplSec	1300	.2015385	.4013036
FemmeEt-tSec	1300	.3115385	.4633005
FemmeDiplC-1	1300	.2261538	.4185009
FemmeDiplUni	1300	.1307692	.3372775
HommeDiplC-1	1300	.2238462	.4169806
HommeDiplUni	1300	.1353846	.3422656
frere1	1300	.51	.5000924
frere2	1300	.3053846	.4607474
frere3	1300	.0784615	.269
privé	1300	.0246154	.1550095
catholique	1300	.1538462	.3609401
autreEcole	1300	.0038462	.0619218
monoparental	1300	.1192308	.3241846
parentsBio	1300	.8292308	.376452

Source : ELNEJ et calculs de l'auteure.

BIBLIOGRAPHIE

- Athletics Ontario. 2013. «Quest for Gold». *Athletics Ontario*. <<http://athleticsontario.ca/quest-for-gold/>>. Consulté le 11 juillet 2013.
- Bailey, Richard. 2006. «Physical Education and Sport in School : A Review of Benefits and Outcome. *Journal of School Health*. En ligne. Vol. 76, no. 8, p. 397-401. <<http://www.bromleyssp.co.uk/FCKfiles/File/FESTIVALS%202010%20-%202011/resources/BaileyPEbenefits.pdf>>. Consulté le 3 juillet 2013.
- BC Athletics. 2011. «General criteria for funding support and selection». In *BC Athletics*. En ligne. <<http://www.bcathletics.org/gensel.htm>>. Consulté le 11 juillet 2013.
- Cameron, Collin A. et Pravin K. Trivedi. 2012. « Linear Panel Models : Basics ». In *Microeconometrics : Methods and applications*. En ligne. p.697-742. <http://public.econ.duke.edu/~vjh3/e262p_07S/readings/Cameron_&_Trivedi_Microeconometrics_Chapters_2.pdf>. Consulté le 14 juillet 2013.
- Susan A. Carlson, Janet E. Fulton, Sarah M. Lee, L. Michele Maynard, David R. Brown, Harold W. Kohl, III, and William H. Dietz. 2007. *Physical Education and Academic Achievement in Elementary School: Data From the Early Childhood Longitudinal Study*. American Journal of Public Health. En ligne. Vol. 98, No. 4, pp. 721-727. <<http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2007.117176>>. Consulté le 18 juillet 2013.
- Deslandes, Rollande et Richard Cloutier. 2005. «Pratiques parentales et réussite scolaire en fonction de la structure familiale et du genre des adolescents.» *Revue française de pédagogie*. En ligne. n° 151, p.61-74. <http://ife.ens-lyon.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP_RF151_5.pdf>. Consulté le 5 juillet 2013.
- Finnie, Ross et Richard Muller. 2008. *The effect of family income, parental education and other background factors on access to post-secondary education in Canada*. En ligne. 38p. <http://higherstrategy.com/mesa/pub/pdf/MESA_Finnie_Mueller.pdf>. Consulté le 5 juillet 2013.

- Institut Fraiser. 2010. (22 octobre). « Bulletin des écoles secondaires du Québec 2010 ». En ligne. <<http://www.fraserinstitute.org/fr/research-news/news/display.aspx?id=16729>>. Consulté le 9 juillet 2013.
- Lassalle, Jean-Yves. 1988. «Le contenu du discours : Les prolongements éducatifs». In *Sport et délinquance*, p. 150-163. Paris : Presses universitaires d'Aix-Marseille.
- Lauzon, Francine. 1990. «L'importance et le rôle des activités corporelles dans l'activité humaine». In *L'éducation psychomotrice : Source d'autonomie et de dynamisme*, p. 7-8. Québec : Presses de l'Université du Québec.
- Merrigan, Philip. 2014. « Analyse statistique des données de panel ». 25 p. Presses de l'Université de Montréal.
- Ministère de l'Éducation, du Loisir et du Sport. 2010 (a). *Programme international pour le suivi des acquis (PISA) 2009 : La performance des jeunes en lecture, en mathématiques et en sciences : Résultats obtenus par les élèves québécois de 15 ans*. En ligne. Québec : Ministère de l'Éducation, du Loisir et du Sport, 67 p. <http://www.mels.gouv.qc.ca/sections/publications/EPEPS/Sanction_etudes/RapportPISA_ProgIntSuiviAcquis2009_f.pdf>. Consulté le 4 mai 2013.
- Ministère de l'Éducation, du Loisir et du Sport. 2010 (b). *Évaluation de l'expérience des participants au programme sport-études*. En ligne. Québec : Ministère de l'Éducation, du Loisir et du Sport, 16p. <http://www.mels.gouv.qc.ca/loisirsport/pdf/publications/sportActPhys/Resume_Rapport_Concilier_Etude-Sport.pdf>. Consulté le 10 mai 2013.
- Ministère de l'Éducation, du Loisir et du Sport. 2012. «Soutien financier aux athlètes». *Ministère de l'Éducation, du Loisir et du Sport*. <http://www.mels.gouv.qc.ca/loisirSport/sporthautniveau/index.asp?page=soutien_athletes>. Consulté le 11 juillet 2013.
- Mulholland, Elizabeth. 2008. «L'Influence du sport : Le rapport sport pur». *Sport Pur*. En ligne. (septembre). p.35-36. <http://www.cces.ca/files/pdfs/TS_report_FR_webdownload.pdf>. Consulté le 4 mai 2013.
- OCDE. 2012. «Éducation des personnes nées dans le pays de résidence de parents immigrés». In *OCDE, Trouver ses marques : Les indicateurs de l'OCDE sur l'intégration des immigrés 2012*. Éditions OCDE. En ligne. 14 p. <<http://dx.doi.org/10.1787/9789264073432-9-fr>>. Consulté le 6 juillet 2013.

- Réseau du sport étudiant du Québec. 2010. «La réussite et la persévérance scolaires par le sport étudiant». *Savoir*. En ligne. Vol. 15, no 4, p.17. <http://rseq.ca/media/26649/la_re_ussite_et_la_perse_ve_rance_scolaire.pdf>. Consulté le 12 mai 2013.
- Sallis, James F. *et al.* 1999. « Effects of Health-Related Physical Education on Academic Achivement : Project SPARK ». *Research Quarterly for Exercise and Sport*. En ligne. Vol.70, no.2, p.127-134 <<http://corepe.pbworks.com/w/file/fetch/48098428/resultssallis.pdf>>. Consulté le 18 juillet 2013.
- Semykina, Anastasia et Jeffrey M. Wooldridge. 2008. *Estimating Panel Data models in the presence of Endogeneity and Selection*. En ligne. 56p. <www.sciencedirect.com/science/article/pii/S0304407610000825>. Consulté le 11 juillet 2013.
- Shephard RJ. 1997. *Curricular physical activity and academic performance*. En ligne. *Pediatric Exercise Science* 1997. Vol.9, p.113-126. <<http://www.humankinetics.com/acucustom/sitename/Documents/DocumentItem/12399.pdf>> . Consulté le 18 juillet 2013.
- Torres-Reyna, Oscar. 2007. *Panel Data Analysis: Fixed and Random Effects (using Stata 10.x)*. Diaporama en ligne. Princeton University. p.9-29. <<http://www.princeton.edu/~otorres/Panel101.pdf>>. Consulté le 15 juillet 2013.
- Wooldridge. Jeffrey. « Econometric Analysis of Cross Section and Panel Data ». 737 p. London, England: The MIT Press. 2002.